Computers and Electronics in Agriculture 192 (2022) 106584

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

An interpretable deep learning approach for calibration transfer among multiple near-infrared instruments

Jie Yang [a,b,c], Juntao Li [d], Jie Hu [d], Wenjun Yang [d], Xiaolei Zhang [a,b], Jinfan Xu [a,b], Youchao Zhang [a,b], Xuan Luo [a,b], K.C. Ting [a,c,e], Tao Lin [a,b], Yibin Ying [a,b,*]

[a] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, Zhejiang 310058, China
[b] Key Laboratory of On Site Processing Equipment for Agricultural Products, Ministry of Agriculture and Rural Affairs, China
[c] International Campus, Zhejiang University, Haining, Zhejiang 314400, China
[d] State Key Laboratory of Animal Nutrition, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China
[e] Department of Agricultural and Biological Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

ARTICLE INFO

Keywords:
Calibration transfer
Deep learning
Interpretation
Reliability
Spectral analysis

ABSTRACT

Spectroscopic techniques have been widely applied in agricultural applications. The development of calibration transfer is promising for the robust analysis of spectral data collected by varying instruments. The reliance on standard samples for standardization approaches remains a critical challenge for on-site applications. In this study, a deep learning approach, named DeepTranSpectra (DTS), is proposed to transfer convolutional neural network models among multiple near-infrared spectrometers with different types. The proposed DTS approach effectively avoids the requirements for standard samples by using labeled samples of slave instruments. The calibration transfer analysis is investigated on a soybean meal and one wheat dataset for predicting moisture and crude protein contents. The developed DTS approach demonstrates improved transfer performance compared with three popular standardization approaches, including piecewise direct standardization (PDS), canonical correlation analysis (CCA), and slope and bias correction (SBC). A feature visualization method is leveraged to interpret the transfer mechanism of the DTS approach. The interpretation results show that the DTS approach refines the model parameters to adapt to slave devices based on critical features of the master calibration. The DTS approach provides advanced reliability under different sample selections in Monte-Carlo cross-validation. The integration of deep learning approaches with calibration transfer analysis facilitates agricultural applications for emerging deep learning-based chemometric analysis.

1. Introduction

Near-infrared spectroscopic technology can assess the chemical information of molecules with the characteristics of fast speed, high accuracy, and non-destruction (Pasquini, 2018). Developing a reliable calibration model is critical to effectively analyze spectral data (Peris-Díaz and Krężel, 2021; Zhang et al., 2021). However, the spectroscopic signals naturally contain instrumental artifacts, leaving the calibrated model invalid for the data collected on different devices (Feudale et al., 2002; Woody et al., 2004). Recalibration is a time-consuming and labor-intensive process, as it usually requires sufficient labeled samples on every individual spectrometer. Given the increasingly varying configurations of spectroscopic devices such as dispersive instruments, Fourier transform tools, and miniaturized spectrometers, developing reliable calibration transfer approaches is critical to meet the demands of on-site spectroscopic detection applications (Liu et al., 2014; Z. Yang et al., 2019).

Standardization methods have been proven useful to confront the challenge of calibration transfer (Blank et al., 1996; Bouveresse et al., 1996; Panchuk et al., 2017; Rehman, 2020). Direct standardization (DS) builds a transfer matrix on all the variables of the data collected by the different devices for the same samples. DS approach has been applied to standardize different types of devices including spectrometers and optical spectroscopic systems (Khaydukova et al., 2017; Panchuk et al., 2017; Surkova et al., 2020). Piecewise direct standardization (PDS) and canonical correlation analysis (CCA) modify the transfer matrix by window segmentation and covariance for improved performance (Bouveresse et al., 1996; Fan et al., 2008). Slope and bias correction (SBC)

* Corresponding author at: College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, Zhejiang 310058, China.
E-mail address: ybying@zju.edu.cn (Y. Ying).

https://doi.org/10.1016/j.compag.2021.106584
Received 31 August 2021; Received in revised form 16 November 2021; Accepted 23 November 2021
Available online 29 November 2021

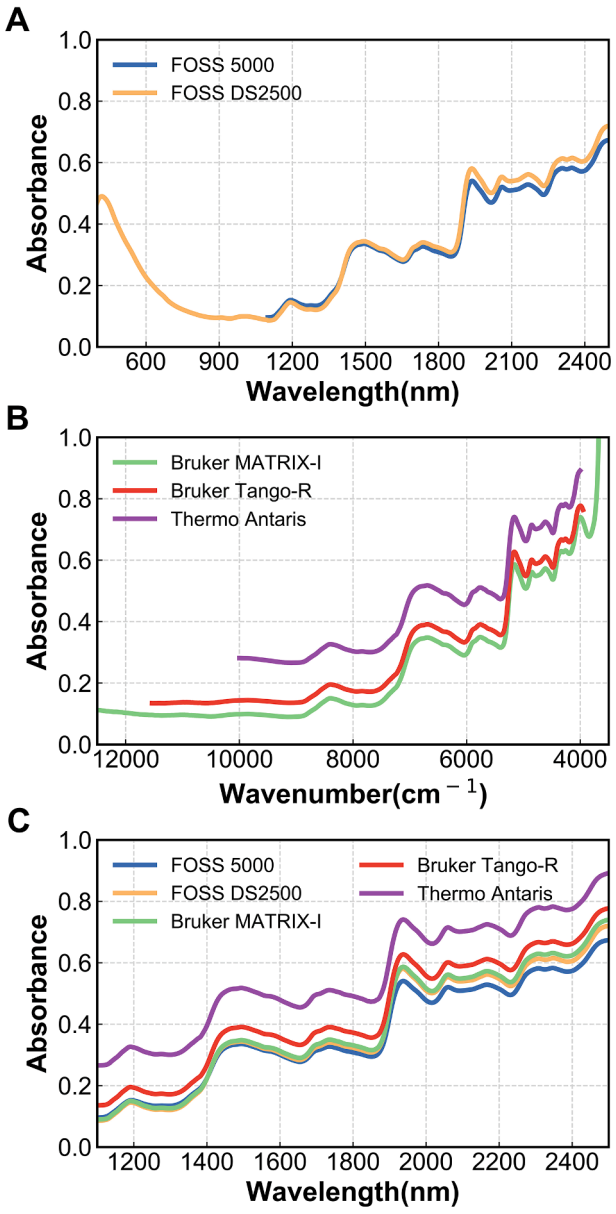


Fig. 1. The band unification of spectral data measured by five different spectrometers. (A) The mean spectra of the soybean meal samples collected by dispersive spectrometers, FOSS 5000 and FOSS DS2500. (B) The mean spectra of the soybean meal samples collected by Fourier transform spectrometers, Bruker MATRIX-I, Bruker Tango-R, and Thermo Antaris. (C) Spectral data processed to the same band range and resolution through wavelength conversion, spline interpolation, and subset selection.

reduces systematic errors by standardizing predictive values (Bouveresse et al., 1996). Existing standard-free calibration transfer methods mainly preprocess data for building robust calibrations (Blank et al., 1996; Woody et al., 2004; Zhao et al., 2019). They are designed to eliminate specific noise but are less effective for unknown variations. The standardization approaches such as PDS, CCA, and SBC still prevail in modern calibration transfer analysis (Bin et al., 2017; Galvan et al., 2020; Liu et al., 2014; Workman, 2018). However, the construction of the transfer matrix is based on standard samples, which are required to be simultaneously measured on different devices.

The reliance on standard samples limits the application of standardization approaches in on-site analytics. The standard sample set usually needs a sufficient sample size with wide variability in the composition for transfer efficacy (Malli et al., 2017; Zhang et al., 2018).

Table 1
The spectral data formats of different instruments.

Instrument	Type	Spectral data range		Resolution	Number of variables
FOSS 5000	Dispersive	1100 nm	2498 nm	2 nm	700
FOSS DS2500	Dispersive	400 nm	2498 nm	2 nm	1050
Bruker MATRIX-I	Fourier transform	12498.51 cm^{-1}	3594.88 cm^{-1}	7.71 cm^{-1}	1154
Bruker Tango-R	Fourier transform	11536 cm^{-1}	3952 cm^{-1}	8 cm^{-1}	949
Thermo Antaris	Fourier transform	9997.17 cm^{-1}	4003.50 cm^{-1}	7.71 cm^{-1}	778

Existing studies prove that unreasonable selection of standard samples deteriorates the performance of standardization methods, which further narrows the application scope for standardization methods (da Silva et al., 2017; Folch-Fortuny et al., 2017). Providing reliable performance under different sample selections is essential for calibration transfer approaches. Therefore, a critical need still exists to develop a reliable approach to reduce the dependence on standard samples for calibration transfer analysis.

Recent advances in artificial intelligence demonstrate that deep learning-based approaches can promote model transferability (Ayres et al., 2021; Devlin et al., 2018; Huynh et al., 2016). Transfer learning methods are primary strategies designed for domain adaptation analysis (Pan and Yang, 2010; Tan et al., 2018). Transfer learning approach is developed to retain learned features from a source task and improve model performance for the target task by sharing domain knowledge, reducing the need for standard samples in model transfer (Padarian et al., 2019). Existing studies have reported that deep convolutional neural network models developed on laboratory bacteria spectroscopy datasets can be effectively extended to clinical detection (Ho et al., 2019; Seddiki et al., 2020). Deep transfer learning methods improved microplastic contamination classification, interregional compositional prediction, and moisture prediction with different conditions in Vis-NIR spectroscopy-based soil analysis (Ng et al., 2020; Padarian et al., 2019; Chen et al., 2021). Recent research has conducted preliminary studies to transfer deep neural networks between two spectrometers (Liu et al., 2018; Mishra and Passos, 2021). However, the interpretation of the mechanism on transfer learning methods remains unexplored. The difference in performance and reliability between transfer learning methods and typical standardization techniques such as PDS, CCA, and SBC needs further investigation. The exploration of model interpretation and reliability is critically needed to ensure the credibility of advanced deep learning techniques for calibration transfer analysis.

This study is designed to evaluate the feasibility of the deep learning-based calibration transfer analysis among multiple near-infrared spectrometers of different types. The working mechanism of the transfer learning strategy to maintain learned features in the model transfer is explored. The research objectives are to (1) develop a deep learning approach for calibration transfer analysis and compare with conventional standardization methods, (2) interpret the mechanism of the proposed deep learning approach by feature visualization, and (3) evaluate the transfer reliability under different sample selections by Monte-Carlo cross-validation.

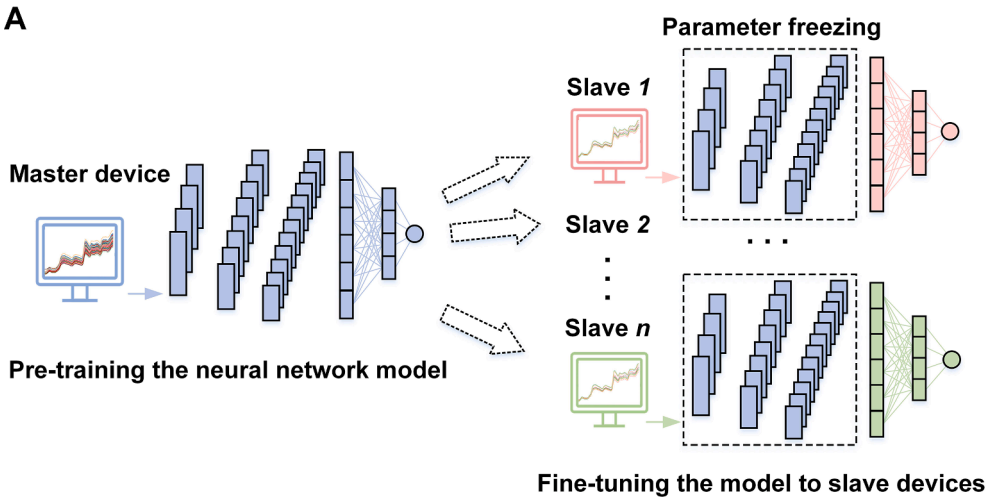
2. Materials and methods

2.1. Data description

The calibration transfer analysis is investigated on a soybean meal dataset and a wheat dataset. All the samples in both datasets are scanned by five different spectrometers, including FOSS 5000, FOSS DS2500, Bruker MATRIX-I, Bruker Tango-R, and Thermo Antaris.

Table 2
The statistics of research compositions in selected datasets.

Dataset	Composition	Sample size	Mean (%)	Max. (%)	Min. (%)	Std.	CV.
Soybean meal	Moisture	171	10.443	12.270	8.805	0.638	0.061
Wheat	Crude protein	172	13.897	17.250	9.920	1.326	0.095



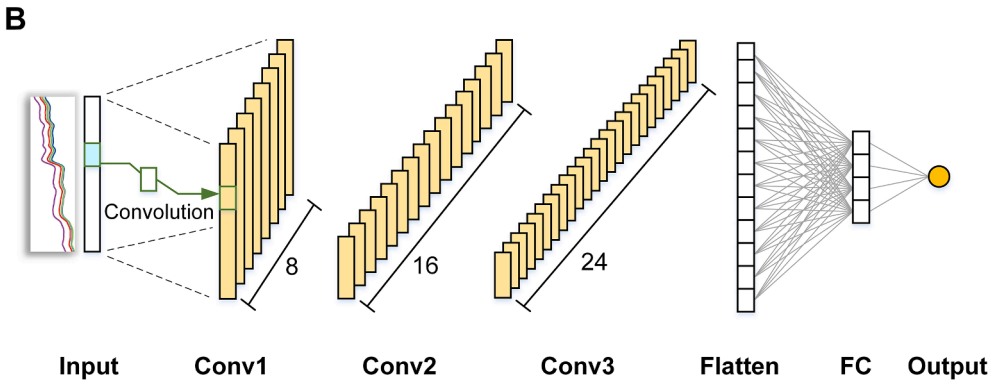


Fig. 2. The illustration on the DTS approach for calibration transfer analysis. (A) The model pre-training and fine-tuning procedures of the DTS approach. The model fine-tuning by the DTS approach does not change the architecture of the pre-trained CNN model. The parameter freezing means the model parameters in convolutional layers are fixed in model fine-tuning. (B) The architectural configuration of the CNN calibration model. The yellow modules represent convolutional layers, and the white modules are the flatten or fully connected layers.

The significant variations in spectral signals between devices are primarily due to the varying optical sensing principles of the dispersive and Fourier transform (FT) spectrometers (Fig. 1A and 1B). The dispersive instruments separate the wavelengths of light by a diffraction grating and direct each wavelength to the detectors, while FT tools collect spectra by an interferometer system. The different spectrometers lead to varying spectral ranges, resolutions, and the number of variables (Table 1).

Spectral data acquired by five different instruments are uniformly processed to the same range of 1100 to 2498 nm with a resolution of 2 nm through wavenumber transformation, spline interpolation, and subset selection. The unit of Fourier transform spectra, wavenumber (cm^{-1}), is converted to the dispersive unit, wavelength (nm), by inversion. The converted spectral data are interpolated to the same resolution by cubic spline interpolation (Bin et al., 2017). The spectral data collected by different devices show significant differences after processing, possibly caused by the baseline shifts and unknown instrumental noise (Fig. 1C).

The moisture content and crude protein content are investigated as the target values of the soybean meal and wheat datasets. Chemical analysis is conducted for the moisture and crude protein, according to the Association of Official Analytical Chemists (AOAC) method 930.15 and 999.03 (Horwitz, 2010). The statistics of selected datasets are shown in Table 2.

2.2. *DeepTranSpectra approach*

The DeepTranSpectra (DTS) approach is proposed to transfer convolutional neural network (CNN) models among multiple different spectrometers based on the transfer learning strategy. The CNN model is derived from the conventional feedforward neural network that converts the stacked fully connected layers to a sparse connection via convolutional kernels (Zhang et al., 2019; Y. Wang et al., 2021). This technique reduces the number of parameters and enables a deeper architecture with multiple processing layers to learn the feature representations from input data hierarchically (LeCun et al., 2015).

The inherent multilayer architecture by the CNN model facilitates the transfer learning for a slave device by model fine-tuning (Fig. 2A). The implementations of the DTS approach are to (1) train a CNN calibration model on the master devices and save all the parameters, (2) load the pre-trained calibration model and freeze the parameters in convolutional layers, and (3) fine-tune the parameters in fully connected layers using the labeled data collected on the slave device. The frozen parameters in the convolutional layers are shared with slave models to maintain the extracted features by the master calibration model. The fine-tuning procedure is to retrain the parameters in fully connected

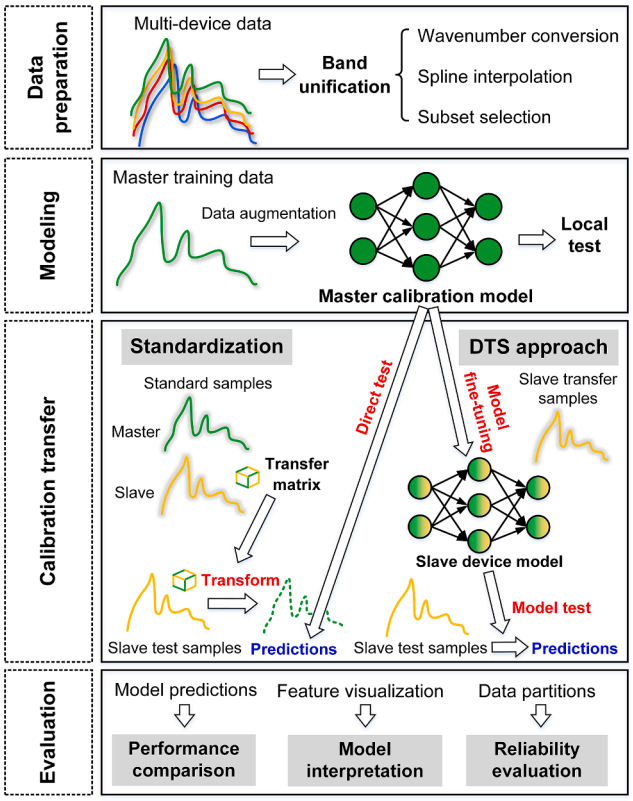


Fig. 3. The calibration transfer workflow for standardization and DTS approaches.

layers, taking the pre-trained weights as the initial values, for adapting to slave devices. Given a data-driven parameter learning mechanism of the neural networks, this approach can preserve general features during twice model training (Padarian et al., 2019). Therefore, the DTS approach can effectively generalize the master calibration model to different slave instruments without standardization.

A typical CNN architecture in spectral analysis is developed as the calibration model for the master device (Liu et al., 2017; J. Yang et al., 2019). The CNN model consists of three convolutional layers, one flatten layer, two fully connected layers including an output layer (Fig. 2B). The numbers of kernels in three convolutional layers are 8, 16, and 24. The batch normalization layer is utilized after each convolutional layer to speed up the training process. Mean squared error (MSE) incorporated with L2 norm regularization, as the objective function, is used to indicate the errors between ground truth and predicted values to guide the model update. Adam optimizer with backpropagation is adopted to optimize the weights of the network. The learning rate scheduler is leveraged to decay the learning rate when the model stops improving. Early stopping is applied to prevent the model from overfitting in the training process. The hyperparameters are decided after searching among preset values based on optimizing the training loss (Mahsereci et al., 2017). The hyperparameter search space and optimized values are provided in Tables A1 and A2.

The architectural configuration of the master calibration model and weights in the convolutional layers are fixed in calibration transfer. The pre-trained weights in fully connected layers are taken as the initial values to retain the general features in the calibration transfer. The hyperparameters related to model fine-tuning are optimized, according

Table 3
The sample sizes of training, transfer, and test set for the calibration transfer analysis.

Dataset	Training set (Augmented)	Transfer set	Test set
Soybean meal	103 (1030)	34	34
Wheat	104 (1040)	34	34

Table 4
The calibration performance on the master device for the soybean meal dataset.

Calibration model	Training RMSE	Training R^2	Test RMSE	Test R^2
CNN	0.095	0.977	0.097	0.977
CNN[a]	**0.079**	**0.984**	**0.078**	**0.985**
PLS	**0.036**	**0.997**	0.083	0.983
PLS[a]	0.043	0.995	**0.076**	**0.986**

[a] The model is built with training data augmentation.

Table 5
The calibration performance on the master device for the wheat dataset.

Calibration model	Training RMSE	Training R^2	Test RMSE	Test R^2
CNN	0.074	0.997	0.145	0.966
CNN[a]	**0.026**	**0.999**	**0.053**	**0.996**
PLS	**0.109**	**0.994**	0.193	0.940
PLS[a]	0.121	0.993	**0.130**	**0.973**

[a] The model is built with training data augmentation.

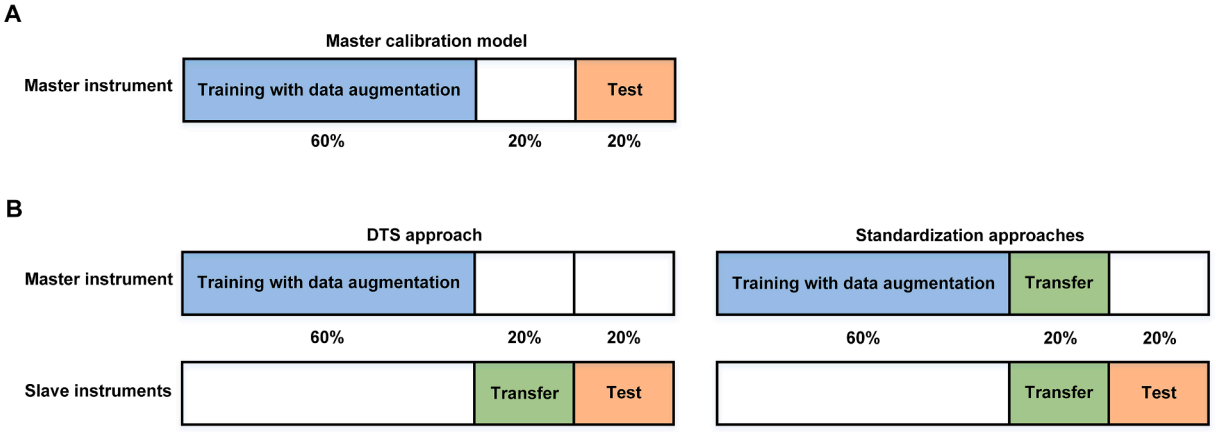


Fig. 4. The schematic of data split for experiments. The aligned block means the sample partition is fixed in developing the master calibration model and calibration transfer analysis. The colored blocks indicate the data utilized in the experiment, while white blocks indicate the data not used. (A) Data collected on the master device is to develop the CNN and PLS calibration models. (B) The master calibration models are transferred to the slave devices by the DTS and standardization approaches.

Table 6
Transfer performance comparison of different approaches on the test samples of the soybean meal dataset.

Transfer approach		FOSS DS2500	Bruker MATRIX-I	Bruker Tango-R	Thermo Antaris
DTS	RMSE	**0.107**	**0.124**	**0.114**	**0.128**
	R^2	**0.972**	**0.963**	**0.968**	**0.960**
CNN-PDS	RMSE	0.116	0.159	0.140	0.180
	R^2	0.967	0.938	0.952	0.921
CNN-CCA	RMSE	0.266	0.304	0.620	0.476
	R^2	0.826	0.773	0.056	0.444
CNN-SBC	RMSE	0.138	**0.124**	0.118	0.617
	R^2	0.950	**0.960**	0.965	0.065
PLS-PDS	RMSE	0.122	0.132	0.141	0.158
	R^2	0.963	0.958	0.951	0.939
PLS-CCA	RMSE	0.260	0.265	0.614	0.251
	R^2	0.834	0.827	0.074	0.846
PLS-SBC	RMSE	0.126	0.144	0.132	0.145
	R^2	0.961	0.949	0.957	0.948
PLS-recalibration	RMSE	0.120	0.153	0.141	0.148
	R^2	0.965	0.942	0.951	0.946

Table 7
Transfer performance comparison of different approaches on the test samples of the wheat dataset.

Transfer approach		FOSS DS2500	Bruker MATRIX-I	Bruker Tango-R	Thermo Antaris
DTS	RMSE	0.098	**0.138**	**0.140**	**0.131**
	R^2	0.985	**0.970**	**0.969**	**0.973**
CNN-PDS	RMSE	**0.080**	0.175	**0.140**	0.160
	R^2	**0.990**	0.951	**0.969**	0.959
CNN-CCA	RMSE	0.224	0.915	0.668	2.781
	R^2	0.919	−0.346	0.282	−11.430
CNN-SBC	RMSE	0.094	0.144	0.164	0.196
	R^2	0.986	0.967	0.957	0.938
PLS-PDS	RMSE	0.152	0.241	0.182	0.174
	R^2	0.963	0.907	0.947	0.951
PLS-CCA	RMSE	0.289	0.726	0.579	2.156
	R^2	0.866	0.154	0.461	−6.474
PLS-SBC	RMSE	0.358	0.174	0.166	0.159
	R^2	0.794	0.951	0.956	0.959
PLS-recalibration	RMSE	0.281	0.306	0.326	0.234
	R^2	0.873	0.849	0.829	0.912

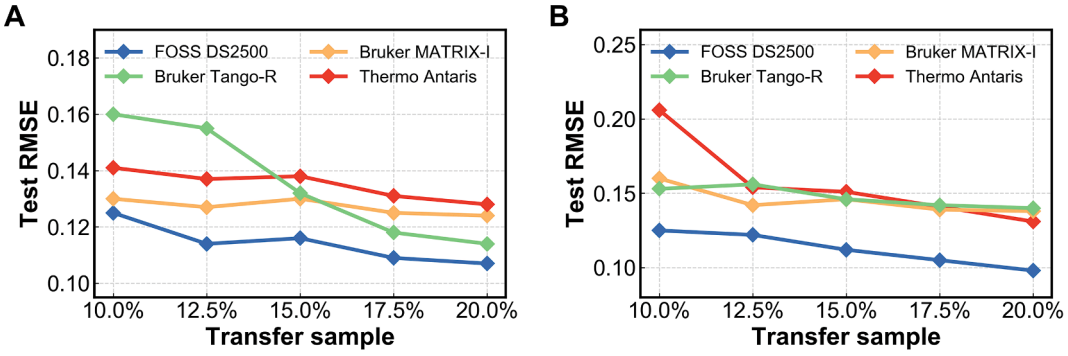


Fig. 5. The impact of different transfer sample sizes on the test performance for the DTS approach. (A) Soybean meal dataset. (B) Wheat dataset.

to the performance of transfer samples, when the model is transferred to a slave instrument. The preset and final parameters are provided in Tables A3 and A4.

2.3. Experimental design

The FOSS 5000 is selected as the master instrument to develop the calibration model. The other four devices act as slave instruments to test the transfer performance. The samples in the soybean meal and wheat datasets are randomly divided into training, transfer, and test set with a ratio of 6:2:2. The workflow of experiments is illustrated in Fig. 3.

2.3.1. Calibration on the master instrument

The master calibration model is developed on the training data collected by the FOSS 5000. Considering there are only about a hundred training samples, the data augmentation technique is leveraged to increase sample size and reduce the risk of overfitting. The training sets are augmented tenfold to 1030 and 1040 samples for soybean meal and wheat datasets, respectively, through adding random variations in slope, multiplication, and offset, as suggested by a previous study (Bjerrum et al., 2017). The test data collected on the master device is to evaluate the calibration performance (Fig. 4A).

The convolutional neural network (CNN) and partial least squares (PLS) models are developed on the master instrument. The architectural

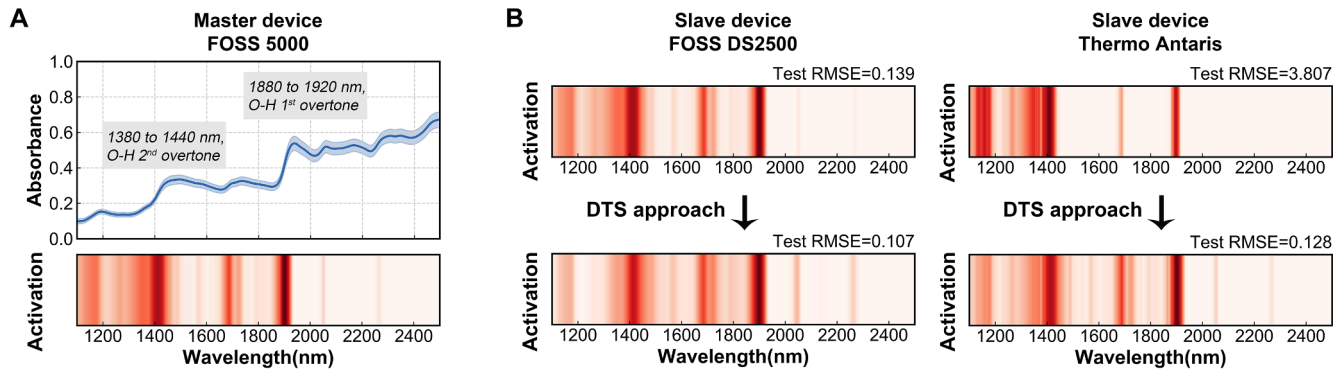


Fig. 6. Feature visualization for the interpretation of the DTS approach. (A) The model activations of the master CNN calibration model. (B) The model activations and predictive accuracies change before and after model fine-tuning by the DTS approach. The upper model activations and accuracies are produced using the master calibration model to directly predict the test samples of slave instruments. The lower model activations and accuracies are produced after calibration transfer by the DTS approach. Two slave devices, FOSS DS2500 and Thermo Antaris, are selected for illustration.

Table 8
P-values of paired two-sample *t*-test based on Test RMSE and Test R^2 for the comparison of DTS to conventional standardization methods.

	Test RMSE	Test R^2
CNN-PDS	<0.01	<0.01
PLS-PDS	<0.01	<0.01
SBC-PDS	<0.01	<0.01

configuration of the CNN model has been provided in the previous section. The number of latent variables (LVs) of the PLS model is optimized in the range of 1 to 20 by 5-fold cross-validation on the training set. The test performances with and without training data augmentation are compared to select a reasonable master calibration model.

2.3.2. *Calibration transfer analysis*

The labeled transfer samples collected on the slave device are required to fine-tune the master calibration model for the DTS approach. The standardization methods need transfer samples (standard samples) to be measured both on the master and slave instruments (Fig. 4B). The test samples of the slave device are to evaluate the calibration transfer performance (Table 3).

The baseline standardization approaches applied in this study include piecewise direct standardization (PDS), canonical correlation analysis (CCA), and slope and bias correction (SBC) (Bouveresse et al., 1996; Fan et al., 2008). The detailed standardization manners of the three methods have been provided in previous studies. The window size of the PDS method is optimized in the preset values of 1, 3, 5, and 7 for an optimal result in each prediction (Chen et al., 2016; Fan et al., 2018). Considering transfer sample sizes are sufficient to develop an independent PLS calibration for slave instruments, the PLS-recalibration method is also selected as a baseline method to compare with the DTS approach. The LV number of the PLS model is optimized in the range of 1 to 20 by 5-fold cross-validation on the transfer set.

The DTS approach is only applicable to neural network models, while the standardization methods can be used for CNN and PLS calibration models. The CNN model calibrated on the master device serves as the base calibration to be transferred by the DTS and three standardization approaches. The master PLS model works as the base calibration for three standardization methods. Therefore, the DTS approach is compared with seven conventional techniques, CNN-PDS, CNN-CCA, CNN-SBC, PLS-PDS, PLS-CCA, PLS-SBC, and PLS-recalibration, to evaluate transfer performance.

To evaluate the impact of transfer sample sizes on the DTS approach, five different transfer sets with 10.0%, 12.5%, 15.0%, 17.5%, 20.0% of all samples are selected for model fine-tuning. The samples with different proportions are randomly chosen from the original transfer set. The test set is the same for a fair comparison of transfer performance.

2.3.3. *Feature visualization for the DTS approach*

The feature visualization analysis is conducted on the soybean meal dataset to interpret how the DTS approach maintains and refines critical features in calibration transfer. The gradient-weighted class activation mapping (Grad-CAM) method has been proven to be a useful tool to capture decisive input variables for the CNN models (Fukuhara et al., 2019; Selvaraju et al., 2017; C.-Y. Wang et al., 2021). The Grad-CAM approach calculates the gradients of the output with respect to the feature maps produced by the last convolutional layer. The gradient values indicate the element-wise importance of the feature maps for the prediction. The summation of all elements' gradients in each feature map reflects the importance score of the current feature map (Selvaraju et al., 2017). The feature maps are weighted summed by the scores to

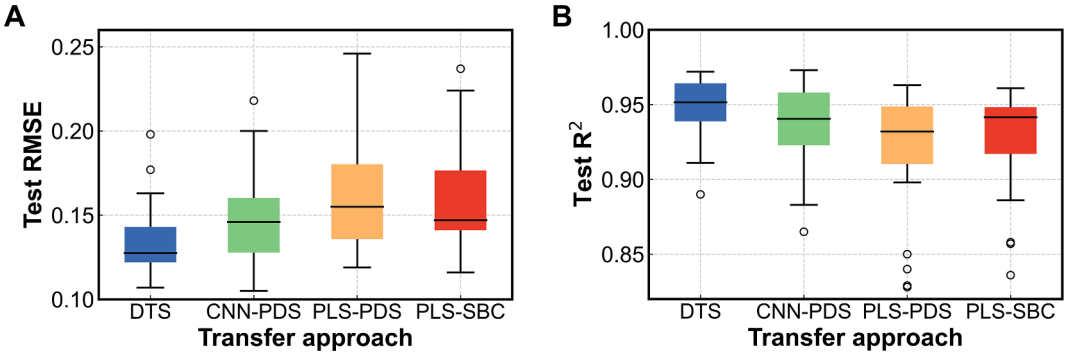


Fig. 7. The evaluation of calibration transfer predictions by selected approaches for different data partitions.

produce a comprehensive response for input variables. The CNN model produces varying feature maps for different input data; thus, the Grad-CAM approach generates sample-specific activation (Fukuhara et al., 2019; C.-Y. Wang et al., 2021). The mean response of test samples is used to reflect an overall model activation.

In this study, the Grad-CAM approach is first applied to visualize the important spectral regions extracted by the master calibration model. The model activations for slave devices, before and after model fine-tuning, are further compared to interpret the model transfer mechanism of the DTS approach.

2.3.4. Reliability evaluation

The reliability of calibration transfer approaches is evaluated on the soybean meal dataset. We randomly produce another five different data partitions for training, transfer, and test set with the ratio of 6:2:2. The overall transfer performance for four slave instruments under all data partitions is comprehensively analyzed to compare the transfer performance of the DTS approach with selected standardization methods. The model parameters are optimized for each data partition.

2.4. Model implementation and evaluation

The performance indicators to evaluate the model are root mean squared error (RMSE) and coefficient of determination (R^2), which are shown in Eqs. (1) and (2).

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}\left[(y_n - \hat{y}_n)^2\right]}{N}} \tag{1}$$

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}{\sum_{n=1}^{N}(y_n - \bar{y})^2} \tag{2}$$

where y_n and $\hat{y}_n$ are the measured and predicted values, respectively. $\bar{y}$ is the mean of y values, and N is the sample size.

All the analytical models were implemented on a Python platform using PyTorch and Scikit-Learn libraries, and MATLAB software with PLS Toolbox. The computations were implemented on a Linux workstation (Ubuntu 16.04 LTS) with an NVIDIA Tesla P40 graphics card with 24 GB of RAM.

3. Results and discussion

3.1. Calibration performance on the master instrument

The CNN and PLS models are developed on the master device, FOSS 5000, for the soybean meal and wheat datasets (Tables 4 and 5). The data augmentation technique improves the test performance of CNN and PLS models for both datasets. The increased variations of training data possibly enhance model robustness for prediction. Therefore, the models developed with data augmentation are selected as the master calibration models for subsequent calibration transfer analysis. The data augmentation slightly deteriorates the training performance of PLS models, which may be caused by the limited data fitting capacity of the linear PLS model (Tables 4 and 5).

For the soybean meal dataset, the CNN and PLS calibration models provide a comparable prediction performance with the test RMSE of 0.078 and 0.076 (Table 4). The CNN model yields test RMSE of 0.053, on the wheat dataset, which significantly outperforms the PLS model with the test RMSE equaled 0.130 (Table 5).

3.2. Transfer performance comparison between the DTS and standardization approaches

The DeepTranSpectra (DTS) approach demonstrates improved transfer performance in terms of a lower test RMSE with an increased test R^2 (coefficient of determination) than standardization approaches.

For the soybean meal dataset, the DTS approach yields the lowest mean RMSE of 0.118 and the highest mean R^2 of 0.966 on four slave devices (Table 6). This approach significantly reduces test RMSE compared with CNN-PDS, PLS-PDS, PLS-SBC, and PLS-recalibration (P-value < 0.05). The DTS approach outperforms all the standardization methods with at least a 14% average reduction of test RMSE on four slave devices. For wheat analysis, overall, the DTS approach exceeds other methods (Table 7). This technique reduces an average test RMSE of 9% compared with the optimal standardization method CNN-PDS.

The transferred models by the DTS approach outperform the calibration models directly developed on the slave devices. The calibration accuracies for four slave instruments by the CNN models are provided in Tables A5 and A6. The DTS approach reduces about 13% and 32% mean test RMSE compared with the slave CNN calibration models for soybean meal and wheat datasets. The transferred models provide test accuracies close to the master calibration models compared with those directly developed on slave instruments. The proposed DTS approach possibly maintains the features of master calibration models in the calibration transfer analysis.

Transfer sample sizes affect the transfer efficacy of the DTS approach. Overall, the performance of the DTS approach is improved with increased transfer samples (Fig. 5). Limited data size, such as 10% transfer samples, significantly deteriorates test accuracy for specific slave instruments, such as Bruker Tango-R for the soybean meal dataset and Thermo Antaris for the wheat dataset. For both datasets, the DTS approach produces acceptable accuracies using 15.0% of all data (about 25 transfer samples), which outperform the calibration models directly developed on the slave instruments (Tables A5 and A6). The performance improves slightly when using over about 30 samples (17.5%). However, the required sample sizes for the DTS approach may be different for varied application scenarios. The DTS approach outperforms CNN-PDS, PLS-PDS, PLS-SBC, and PLS-recalibration under different transfer sample sizes on both datasets (Figure A1).

The advanced transfer performance by the DTS approach is primarily because this approach integrates the CNN calibration model with the model fine-tuning strategy. The multilayer CNN models hierarchically learn the feature representations from input data. Layer-by-layer feature extraction mechanism captures general low-level features in the first few layers with high-level specifics in the deeper layers (LeCun et al., 2015; Zhang et al., 2020). Thus, the frozen parameters in convolutional layers maintain the feature extraction capacity of the master calibration model. The model fine-tuning on deeper layers refines the weights of the extracted features and adapts to slave devices. Therefore, the DTS approach provides satisfactory prediction accuracy for the test samples collected on the slave instruments.

The working principle of DTS approach is highly different from standardization methods. Standardization techniques, such as PDS and CCA, are to numerically remove spectral differences across the entire spectra among varying instruments. The SBC method is to eliminate systematic predictive errors between calibration models (Malli et al., 2017). These approaches are not always effective to different instrumental variations, making it challenging to provide stable transfer performance in different tasks. In this study, the CNN-PDS, PLS-PDS, and PLS-SBC provide acceptable performance for the soybean meal datasets, while CNN-PDS shows superiority among standardization tools for the wheat dataset.

The effectiveness of PDS standardization is mainly due to improved fitting capacity by optimizing window sizes for spectral segmentation. Inappropriate window sizes significantly drop the accuracy of the PLS-PDS approach (Tables A7 and A8). The CNN models are better integrated with the PDS standardization method, which possibly due to the feature extraction of CNN, through kernel sliding, can be well agreed with the window segmentation of PDS. The CCA could not show superior results in the selected datasets. A previous study has proven CCA method might be suitable for datasets with larger sample sizes (Fan et al., 2008).

The PLS-recalibration method yields acceptable accuracies in the soybean meal dataset but not in the wheat dataset. This result may be because PLS-recalibration cannot take advantage of data collected on the master instrument and deteriorates performance without sufficient representative training samples.

3.3. Feature visualization for the DTS approach

The important spectral variables determined by the master CNN calibration model are overall consistent with the theoretical absorption peak of the moisture (Fig. 6A). The model activations primarily yield at around 1880 to 1920 nm and 1380 to 1440 nm, which correspond to the first and second overtones of the O—H bond, respectively (Tres et al., 2012). The result demonstrates the CNN model effectively extracts critical variables from high dimensional spectral variables. However, activations around 1680 to 1700 nm and 1140 to 1180 nm mainly correspond to the C—H bonds independent of water content (Daszykowski et al., 2008; Tres et al., 2012). These spectral variables are possibly related to carbon-containing compositions such as protein and fiber, which potentially affect moisture proportion.

The model activations are normalized importance scores for all input variables to evaluate critical spectral regions learned by the model. The activation values are determined based on the gradients and input variables (Fukuhara et al., 2019; C.-Y. Wang et al., 2021). Therefore, changes in input data usually result in varying activations produced by the same model, which are different from the coefficients of the linear regression models.

The predictive accuracies and model activations change, for the same test samples of slave devices, before and after calibration transfer by model fine-tuning (Fig. 6B). The master calibration model provides acceptable performance with RMSE equaled 0.139 for the test set of FOSS DS2500. The model activations of the master calibration remain in O—H-related spectral regions to maintain the modeling features. This is because that spectral data collected by FOSS DS2500 show similarity with that of the master device FOSS 5000 (Fig. 1C). The fine-tuning approach enables the model to further refines parameters to adapt to slave device with the RMSE reduced to 0.107. For test samples of Thermo Antaris, the model activations show significant differences. The importance scores around 1410 nm and 1160 nm rise, and responses around 1900 nm decrease. The master calibration model provides deteriorated accuracy with an RMSE of 3.807. The model fine-tuning adjusts the parameters in fully connected layers and therefore reorganizes the model activations for adapting to the slave devices. The transferred model reduces the test RMSE to 0.128 and shows similar activations compared with that of the master calibration model. In conclusion, the feature visualization results interpret that the model fine-tuning by the DTS approach effectively maintains critical features of the master calibration model and adjusts the model for adapting slave devices to provide satisfactory calibration transfer performance.

3.4. Reliability evaluation by Monte-Carlo cross-validation

The DTS approach significantly improves transfer reliability under different data partitions for slave instruments compared with selected standardization methods (Table 8). The DTS approach produces the lowest mean RMSE with the smallest interquartile range and extremum (Fig. 7). The CNN-PDS method shows superiority among standardization methods and yields comparable performance on Bruker MATRIX-I and Bruker Tango-R (Table A9).

The results demonstrate the proposed DTS approach provides reliable performance and is less impacted by different sample selections. The DTS approach can be a viable option for practical calibration transfer applications.

4. Conclusions

A deep learning approach, named DeepTranSpectra (DTS), is developed to transfer the convolutional neural network models to multiple near-infrared spectrometers with different types. The proposed approach successfully avoids the requirements of standard samples measured on various instruments by utilizing labeled data collected on the slave devices. The results show the DTS approach improves transfer performance compared with conventional standardization methods, including PDS, SBC, and CCA, on the soybean meal and wheat datasets. The feature visualization analysis demonstrates that the proposed approach effectively maintains critical features of the master calibration model and adapts the model to slave instruments. The DTS approach provides advanced reliability than standardization methods under different sample selections in Monte-Carlo cross-validation.

The proposed DTS approach possibly facilitates the practical application of deep learning-based spectral analysis. Further study could investigate the feasibility of this approach to the calibration transfer analysis of other analytical devices and sensing systems in agriculture.

CRediT authorship contribution statement

Jie Yang: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Juntao Li:** Investigation, Writing – review & editing. **Jie Hu:** Investigation. **Wenjun Yang:** Investigation. **Xiaolei Zhang:** Methodology, Software, Writing – review & editing. **Jinfan Xu:** Methodology, Software. **Youchao Zhang:** Methodology, Software. **Xuan Luo:** Writing – review & editing. **K.C. Ting:** Supervision, Writing – review & editing. **Tao Lin:** Conceptualization, Supervision, Writing – review & editing. **Yibin Ying:** Supervision, Funding acquisition.

Declaration of Competing Interest

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.

Acknowledgments

This work was partially funded by the Key Joint Fund for Regional Innovation Development (U20A2019), the National Natural Science Foundation of China under Grant Number F030601, and Zhejiang University.

Code availability

The code that supports the findings of this study is available from the corresponding author upon reasonable request.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2021.106584.

References

Ayres, L.B., Gomez, F.J., Linton, J.R., Silva, M.F., Garcia, C.D., 2021. Taking the leap between analytical chemistry and artificial intelligence: A tutorial review. Anal. Chim. Acta 338403.

Bin, J., Li, X., Fan, W., Zhou, J.-H., Wang, C.-W., 2017. Calibration transfer of near-infrared spectroscopy by canonical correlation analysis coupled with wavelet transform. Analyst 142 (12), 2229–2238.

Bjerrum, E.J., Glahder, M., Skov, T., 2017. Data augmentation of spectral data for convolutional neural network (CNN) based deep chemometrics. ArXiv Prepr. ArXiv171001927.

Blank, T.B., Sum, S.T., Brown, S.D., Monfre, S.L., 1996. Transfer of near-infrared multivariate calibrations without standards. Anal. Chem. 68 (17), 2987–2995.

Bouveresse, E., Hartmann, C., Massart, D.L., Last, I.R., Prebble, K.A., 1996. Standardization of near-infrared spectrometric instruments. Anal. Chem. 68 (6), 982–990.
Chen, W.-R., Bin, J., Lu, H.-M., Zhang, Z.-M., Liang, Y.-Z., 2016. Calibration transfer via an extreme learning machine auto-encoder. Analyst 141 (6), 1973–1980.
Chen, Y.u., Li, L., Whiting, M., Chen, F., Sun, Z., Song, K., Wang, Q., 2021. Convolutional neural network model for soil moisture prediction and its transferability analysis based on laboratory Vis-NIR spectral data. Int. J. Appl. Earth Obs. Geoinformation 104, 102550. https://doi.org/10.1016/j.jag.2021.102550.
da Silva, N.C., Cavalcanti, C.J., Honorato, F.A., Amigo, J.M., Pimentel, M.F., 2017. Standardization from a benchtop to a handheld NIR spectrometer using mathematically mixed NIR spectra to determine fuel quality parameters. Anal. Chim. Acta 954, 32–42.
Daszykowski, M., Wrobel, M.S., Czarnik-Matusewicz, H., Walczak, B., 2008. Near-infrared reflectance spectroscopy and multivariate calibration techniques applied to modelling the crude protein, fibre and fat content in rapeseed meal. Analyst 133 (11), 1523. https://doi.org/10.1039/b803687j.
Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. ArXiv Prepr. ArXiv181004805.
Fan, W., Liang, Y., Yuan, D., Wang, J., 2008. Calibration model transfer for near-infrared spectra based on canonical correlation analysis. Anal. Chim. Acta 623 (1), 22–29.
Fan, X., Lu, H., Zhang, Z., 2018. Direct calibration transfer to principal components via canonical correlation analysis. Chemom. Intell. Lab. Syst. 181, 21–28.
Feudale, R.N., Woody, N.A., Tan, H., Myles, A.J., Brown, S.D., Ferré, J., 2002. Transfer of multivariate calibration models: a review. Chemom. Intell. Lab. Syst. 64 (2), 181–192.
Folch-Fortuny, A., Vitale, R., de Noord, O.E., Ferrer, A., 2017. Calibration transfer between NIR spectrometers: New proposals and a comparative study. J. Chemom. 31 (3), e2874. https://doi.org/10.1002/cem.2874.
Fukuhara, M., Fujiwara, K., Maruyama, Y., Itoh, H., 2019. Feature visualization of Raman spectrum analysis with deep convolutional neural network. Anal. Chim. Acta 1087, 11–19.
Galvan, D., Bona, E., Borsato, D., Danieli, E., Montazzolli Killner, M.H., 2020. Calibration transfer of partial least squares regression models between desktop nuclear magnetic resonance spectrometers. Anal. Chem. 92 (19), 12809–12816.
Ho, C.-S., Jean, N., Hogan, C.A., Blackmon, L., Jeffrey, S.S., Holodniy, M., Banaei, N., Saleh, A.A., Ermon, S., Dionne, J., 2019. Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. Nat. Commun. 10, 1–8.
Horwitz, W., 2010. Official methods of analysis of AOAC International. In: Horwitz, William (Ed.), Agricultural chemicals, contaminants, drugs, vol. I. AOAC International, Gaithersburg (Maryland), p. 1997.
Huynh, B.Q., Li, H., Giger, M.L., 2016. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J. Med. Imaging 3 (3), 034501. https://doi.org/10.1117/1.JMI.3.3.034501.
Khaydukova, M., Medina-Plaza, C., Rodriguez-Mendez, M.L., Panchuk, V., Kirsanov, D., Legin, A., 2017. Multivariate calibration transfer between two different types of multisensor systems. Sens. Actuators B Chem. 246, 994–1000.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Liu, J., Osadchy, M., Ashton, L., Foster, M., Solomon, C.J., Gibson, S.J., 2017. Deep convolutional neural networks for Raman spectrum recognition: a unified solution. Analyst 142 (21), 4067–4074.
Liu, L., Ji, M., Buchroithner, M., 2018. Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery. Sensors 18 (9), 3169. https://doi.org/10.3390/s18093169.
Liu, Y., Cai, W., Shao, X., 2014. Standardization of near infrared spectra measured on multi-instrument. Anal. Chim. Acta 836, 18–23.
Mahsereci, M., Balles, L., Lassner, C., Hennig, P., 2017. Early Stopping without a Validation Set. ArXiv170309580 Cs Stat.
Malli, B., Birlutiu, A., Natschläger, T., 2017. Standard-free calibration transfer-An evaluation of different techniques. Chemom. Intell. Lab. Syst. 161, 49–60.
Mishra, P., Passos, D., 2021. Deep calibration transfer: Transferring deep learning models between infrared spectroscopy instruments. Infrared Phys. Technol. 103863.
Ng, W., Minasny, B., McBratney, A., 2020. Convolutional neural network for soil microplastic contamination screening using infrared spectroscopy. Sci. Total Environ. 702, 134723. https://doi.org/10.1016/j.scitotenv.2019.134723.
Padarian, J., Minasny, B., McBratney, A.B., 2019. Transfer learning to localise a continental soil vis-NIR calibration model. Geoderma 340, 279–288.
Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359.
Panchuk, V., Kirsanov, D., Oleneva, E., Semenov, V., Legin, A., 2017. Calibration transfer between different analytical methods. Talanta 170, 457–463.
Pasquini, C., 2018. Near infrared spectroscopy: A mature analytical technique with new perspectives–A review. Anal. Chim. Acta 1026, 8–36.
Peris-Díaz, M.D., Krężel, A., 2021. A guide to good practice in chemometric methods for vibrational spectroscopy, electrochemistry, and hyphenated mass spectrometry. TrAC Trends Anal. Chem. 135, 116157. https://doi.org/10.1016/j.trac.2020.116157.
Rehman, T.U., 2020. Calibration transfer across multiple hyperspectral imaging-based plant phenotyping systems_ I - Spectral space adjustment. Comput. Electron, Agric, p. 11.
Seddiki, K., Saudemont, P., Precioso, F., Ogrinc, N., Wisztorski, M., Salzet, M., Fournier, I., Droit, A., 2020. Cumulative learning enables convolutional neural network representations for small mass spectrometry data classification. Nat. Commun. 11, 1–11.
Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: in: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626.
Surkova, A., Bogomolov, A., Legin, A., Kirsanov, D., 2020. Calibration transfer for LED-based optical multisensor systems. ACS Sens. 5 (8), 2587–2595.
Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., 2018. A survey on deep transfer learning. In: International Conference on Artificial Neural Networks. Springer, pp. 270–279.
Tres, A., van der Veer, G., Perez-Marin, M.D., van Ruth, S.M., Garrido-Varo, A., 2012. Authentication of organic feed by near-infrared spectroscopy combined with chemometrics: A feasibility study. J. Agric. Food Chem. 60 (33), 8129–8133.
Wang, C.-Y., Ko, T.-S., Hsu, C.-C., 2021a. Interpreting convolutional neural network for real-time volatile organic compounds detection and classification using optical emission spectroscopy of plasma. Anal. Chim. Acta 1179, 338822. https://doi.org/10.1016/j.aca.2021.338822.
Wang, Y., Li, M., Ji, R., Wang, M., Zheng, L., 2021b. A deep learning-based method for screening soil total nitrogen characteristic wavelengths. Comput. Electron. Agric. 187, 106228. https://doi.org/10.1016/j.compag.2021.106228.
Woody, N.A., Feudale, R.N., Myles, A.J., Brown, S.D., 2004. Transfer of multivariate calibrations between four near-infrared spectrometers using orthogonal signal correction. Anal. Chem. 76 (9), 2595–2600.
Workman, J.J., 2018. A review of calibration transfer practices and instrument differences in spectroscopy. Appl. Spectrosc. 72 (3), 340–365.
Yang, J., Xu, J., Zhang, X., Wu, C., Lin, T., Ying, Y., 2019a. Deep learning for vibrational spectral analysis: Recent progress and a practical guide. Anal. Chim. Acta 1081, 6–17.
Yang, Z., Albrow-Owen, T., Cui, H., Alexander-Webber, J., Gu, F., Wang, X., Wu, T.-C., Zhuge, M., Williams, C., Wang, P., Zayats, A.V., Cai, W., Dai, L., Hofmann, S., Overend, M., Tong, L., Yang, Q., Sun, Z., Hasan, T., 2019b. Single-nanowire spectrometers. Science 365 (6457), 1017–1020.
Zhang, F., Zhang, R., Ge, J., Chen, W., Yang, W., Du, Y., 2018. Calibration transfer based on the weight matrix (CTWM) of PLS for near infrared (NIR) spectral analysis. Anal. Methods 10 (18), 2169–2179.
Zhang, X., Lin, T., Xu, J., Luo, X., Ying, Y., 2019. DeepSpectra: An end-to-end deep learning approach for quantitative spectral analysis. Anal. Chim. Acta 1058, 48–57.
Zhang, X., Xu, J., Yang, J., Chen, L.i., Zhou, H., Liu, X., Li, H., Lin, T., Ying, Y., 2020. Understanding the learning mechanism of convolutional neural networks in spectral analysis. Anal. Chim. Acta 1119, 41–51.
Zhang, X., Yang, J., Lin, T., Ying, Y., 2021. Food and agro-product quality evaluation based on spectroscopy and deep learning: A review. Trends Food Sci. Technol. 112, 431–441.
Zhao, Y., Zhao, Z., Shan, P., Peng, S., Yu, J., Gao, S., 2019. Calibration transfer based on affine invariance for NIR without transfer standards. Molecules 24 (9), 1802. https://doi.org/10.3390/molecules24091802.