Infrared Physics & Technology 117 (2021) 103863

Contents lists available at ScienceDirect

Infrared Physics and Technology

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

Deep calibration transfer: Transferring deep learning models between infrared spectroscopy instruments

Puneet Mishra [a,*], Dário Passos [b]

[a] Wageningen Food and Biobased Research, Bornse Weilanden, 9, P.O. Box 17, Wageningen 6700AA, the Netherlands
[b] CEOT, Physics Department, Universidade do Algarve, Campus de Gambelas, FCT Ed.2, Faro 8005-189, Portugal

ARTICLE INFO

Keywords:
Standard-free
Spectroscopy
Model update
Convolutional neural networks

ABSTRACT

Calibration transfer (CT) is required when a model developed on one instrument needs to be transferred and used on a new instrument. Several methods are available in the chemometrics domain to transfer the multivariate calibrations developed using modelling techniques such as partial least-square regression. However, recently deep learning (DL) models are gaining popularity to model spectral data. The traditional multivariate CT methods are not suitable to transfer a deep learning model which is based on neural networks architectures. Hence, this study presents the concept of deep calibration transfer (CT) for transferring a DL model made on one instrument onto a new instrument. The deep CT is based on the concept of transfer learning from the DL domain. To show it, two different CT cases are presented. The first case is the CT between benchtop FT-NIR (Fourier Transform Near Infrared) instruments, and the second case is the CT between handheld NIR (Near Infrared) instruments. In both the demonstrated cases, the transfer was performed standard-free i.e., no common standard samples were used to estimate any transfer function. The results showed that with deep CT, the DL models made on one instrument can be easily adapted and transferred to a new instrument. The main benefit of the deep CT is that it is a standard free approach and does not require any standard sample measurements. Such a standard free approach to transfer DL models between instruments can support a widespread sharing of chemometric DL models between the scientific practitioners.

1. Introduction

Chemometrics calibrations are of high-value, as they are the results of extensive sensor measurements and wet chemistry destructive reference analysis [1]. Furthermore, the development of primary models requires extensive optimisation tasks such as the search for best pre-processing combination, reduction of external influences, optimisation of latent variables for partial least-square (PLS) based models and many more associated data analysis tasks [2–4]. Hence, the user always expects that once the calibration is set up, the need for repeated wet chemistry and data analysis can be bypassed [5]. To a major extent, the expectations of the user are met, however, there are certain situations where the user may need to repeat the wet chemistry and data analysis. For example, when the user decides to increase its research facilities by adding on a new similar instrument or when the user needs to replace an old/damaged instrument with a new one [6,7]. In that case, to be able to calibrate the new instrument, the user may need to repeat the calibration procedure previously done for the old instrument [6–8]. Nonetheless, this calibration step of the new instrument can be skipped if the user already has a trained model that relates the signal from the instrument to the property of interest. The user only needs to account for the differences between the instruments and then the calibration model from the old instrument can be used on the new one [8–12]. In the chemometrics domain, the procedure of modelling and remediate instrument differences to allow an existing model to be used on a new instrument is called calibration transfer (CT) [10,11,13,14].

Different instruments have intrinsic differences, due to their diverse technical components (e.g., detector type, illumination source, difference in spectral range, resolution, variable axis registry, etc.) or due to the surrounding environment in which the instrument is deployed [10,11]. If the instruments for which an existing model needs to be transferred to is identical to the old instrument, the CT task can be understood as easier, i.e., identical instruments will have less differences to compensate [10,15]. However, when the model must be transferred to a non-identical instrument, e.g., from a laboratory benchtop spectrometer to a hand-held consumer spectrometer, then the transfer procedure

* Corresponding author.
E-mail address: puneet.mishra@wur.nl (P. Mishra).

https://doi.org/10.1016/j.infrared.2021.103863
Received 27 May 2021; Received in revised form 30 July 2021; Accepted 1 August 2021
Available online 4 August 2021
1350-4495/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

requires more efforts. In the domain of chemometrics, several approaches to remove the differences between instruments (a.k.a. CT) are available. For example, two of the most widely used CT techniques are direct standardization (DS) [16] and piece direct standardization (PDS) [17]. Both techniques require stable standard samples to be measured on the old and the new instruments, such that the difference between the responses from the two instruments can be modelled with a transfer function [10]. Once the transfer function is modelled, the data from different instruments are transformed by it and the model made on the primary instruments can be used on the secondary instruments [16,17]. Both methods, DS and PDS, work well in most of the scenarios but have as the main limitation that they require some standard samples to be measured on both instruments [11]. Hence, this is not desirable in the scenario when the old instrument is not available (e.g., damaged) or is based in a far location [6,7,11]. In that case, a standard-free chemometric method can be used as it only needs some measurements from the new instrument to remove the difference between instruments [6,7,10].

Several standard-based [10] and standard free CT methods are available in the chemometrics domain [6,7,11]. However, classical CTs approaches were developed relying on the linear nature of classical multivariate models, hence, are only suitable for linear multivariate models, such as PLS (Partial Least Square) regression analysis. Their application to non-linear algorithms such as Support Vector Machines (with RBF kernels) and deep neural networks is not straight forward.

Modern deep learning (DL) algorithms have been slowly gaining traction in the chemometrics community as a powerful tool to model multivariate signals [18–21]. For several tasks/problems, DL models have already outperformed classical chemometric algorithms (e.g., PLS, PCA, etc) [19–21]. Like the classical chemometrics methods, most DL models are also dependent on the instrument characteristics and require adaption before being used on a new instrument. Due to the underlying complex architecture of DL models, CT methods classically available in the chemometrics domain are not suitable. Despite this fact, DL models such as convolutional neural networks (CNNs) also have the potential to be adapted to a new scenario. This can be done by implementing the concept of Transfer Learning (TL) for neural network layers. In the broader area of Machine Learning research, TL has a much wider definition [22–24] that applies to the general transfer of information between different domains. In this work, the concept of TL used is restricted to its applicability within the DL framework, i.e., to what needs to be done so that an already trained DL model can be used as the starting point for a new model in order to apply it to new data observations. More specifically the concept of TL is here adapted for solving the challenges of CT for DL models applied to NIR spectral data from different instruments [25].

Transfer learning in the domain of DL is prevalent in computer vision research [26–30,35]. In this context, the main aim of TL is to use the already acquired knowledge to avoid the expensive task of training large DL models from scratch [23]. In computer vision, TL is particularly useful for CNNs architectures [27–29]. These algorithms have usually two main parts: a convolutional block (usually a combination of convolutional (conv.) and pooling layers) that allows feature extraction and combination, and a mapping block (of Fully Connected layers, FC) that maps the extracted features to the target variable [31]. For certain computer vision tasks, there are evidence that the trained convolutional block (conv. filters) doesn't need retraining when used on the new case (e.g. new data) and only the model's weights in the mapping block (FC layers) need to be adjusted [23]. The ability to find certain visual patterns in an image despite their location is called translational invariance and is a property of deep CNN architectures developed in recent years. However, this property is dependent of the CNN architecture and its ability to merge low level features (from the initial conv. layers) into higher level abstractions (last conv. layers) that lead to the identification of individual patterns in an image, whatever their location. For the case of multivariate signals, such as spectral data, this has not yet been proven to hold true, especially in the case of shallow (just a few layers) CNNs. In spectra, most of the times, the relevant spectral information for the prediction of a certain property (target variable) is sparsely distributed over the whole spectrum and it is not clear that the use of windowed type convolutional filters has the same result as in spatial data (e.g., image). Moreover, spectra acquired with different instruments carry information related not only to the substance being analysed but also spectral signatures that are specific of the instrument (e.g. local shifts in peaks). Taking this information in consideration, for spectral data, retraining the conv. block might provide some extra insights. In recent years, for soil spectral data processing, the use of TL was proposed to update DL models covering added information variability in the data samples [25,32]. More recently, TL was also applied to the problems of new seasonal variability in fruit spectroscopy and changes in ingredients in a melamine manufacturing process [33]. However, none of these works covered the CT problem.

The aim of this study is to show that it is possible to apply the DL transfer learning methodology to the specific problem of transferring DL models developed on data from one instrument to be used on a new instrument. In this work we refer to this method as deep Calibration Transfer (deep CT). To showcase the applicability of the method, a 1D-CNN model was transferred between instruments in two different calibration transfers scenarios: CT between similar benchtop FT-NIR instruments, and CT between similar handheld NIR instruments. The transfer was performed standard-free i.e., no common standard samples were used to estimate any transfer function. In both experiments, the DL model was first developed for the primary instrument and later transferred to the second instrument using deep CT. A comparison between the performance of the transferred models is done by investigating the effect of applying TL to different layer blocks. The present work iterates on [33] and, for the best of our bibliographic search, is the first to explore how and which type of TL can be used for CT applications to transfer spectral DL models between different instruments. A general abstract concept of deep CT is shown in Fig. 1. A key point to note is that this study does not provide any comparison with existing CT techniques in the chemometrics domain because the traditional CT techniques are designed for transferring PLS type models and not deep neural networks.

2. Materials and method

2.1. Data sets

2.1.1. Transfer between bench top instruments – Tablet data set

To show the single response transfer case between laboratory-based benchtop instruments, a publicly available tablet data set was used. This tablet data set is a benchmark[1] data set for testing and evaluating novel calibration model maintenance methods. In this study, the data was obtained from the website: https://eigenvector.com/resources/data-sets/. The tablet data consists of 655 pharmaceutical tablet samples (155 calibration samples, 460 test samples and 40 validation samples) measured on two Multitab spectrometers (Foss-NIRSystems, Silverspring, MD) in the transmittance mode from 600 to 1898 nm in 2 nm increments. Furthermore, as a reference, the assay value of the active ingredient (mg) was used. Based on earlier studies on the same data set, the outliers were removed from the calibration and test set as suggested in [25]. Due to the high signal noise above 1740 nm, data points above this wavelength were discarded in this analysis. The final data set has 642 spectra (600–1740 nm at 2 nm interval) for each instrument 1 and 2 and 642 corresponding assay values of the active ingredient. The range of the assay value was from 151.6 to 239.1 mg.

[1] Information still available through web.archive.org on http://web.archive.org/web/20060312182337/http://www.idrc-chambersburg.org/shootout_2002.htm.

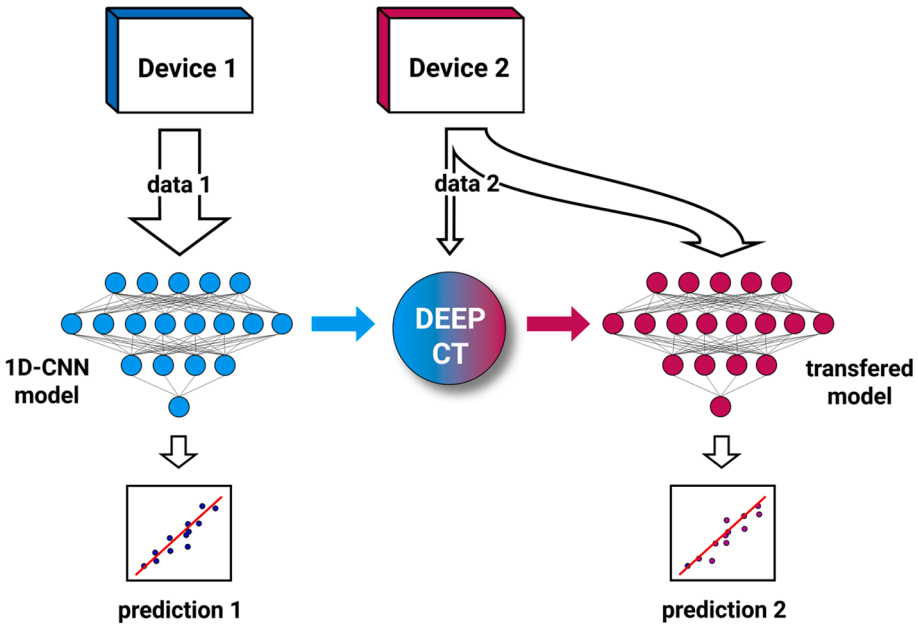

Fig. 1. The concept of deep CT which can be used to transfer deep learning models between instruments.

2.1.2. Transfer between handheld instruments – Olive data set

For the case of calibration transfer between handheld instruments, a NIR data set related to dry matter (DM) prediction in olive fruit was used. The data set was the same as presented in [34] and consists of a total of 583 spectra and reference DM measurements. The spectral measurements were performed with two identical portable spectrometers (Felix F-750, Camas, WA, USA) in the range of 310–1135 nm, with a spectral resolution of 8–13 nm. In this study, the analysis was restricted to the spectral range (750–999 nm) because it contains most of the information about the chemical overtones for moisture (OH) which is inversely related to the DM. The original data had several outliers that were removed using a PLS decomposition and an analysis of T^2 and Q statistics, as described in [33]. After the outlier removal, the total spectra and reference dry matter for instrument 1 were 466, and for instrument 2 were 468. The DM ranged between 23.06 and 37.19 %.

2.2. Data augmentation

In both data sets, the number of samples is relatively small for DL purposes. Hence, to increase the data sets size, a data augmentation procedure was performed according to [36]. This method consists in adding random variations in offset, multiplication and slope of the existing spectra (the target value is maintained). It basically simulates slightly different spectra acquisition scenarios (e.g., light continuum contamination, etc.) so that for the same target variable value, there are now multiple (slightly different) copies of the initial spectra. Offset was varied $\pm$ 0.10 times the standard deviation of the training set, multiplication was done with 1 $\pm$ 0.10 times the standard deviation of the training set, and the slope was adjusted uniformly randomly between 0.95 and 1.05 as described in [36]. The process was repeated 10 times. This data augmentation served two purposes, first it increased the data size to become more suitable for DL modelling and second, it introduced extra variation in the data such that, models trained on it, can become more robust to unseen variations.

In the case of the tablet data set, the data is already partitioned into training, validation, and test sets for both instruments when downloaded from the source. Training and validation subsets are augmented separately, while the test subsets are maintained untouched. A primary DL model is created for the using data from instrument 1 only and then transferred and tested on data from instrument 2. For the model transferring operation, the training and validation data from the instrument 2 were renamed as “fine-tune training” and “fine-tune validation” sets, respectively for a clear distinction. The fine-tune training and fine-tune validation sets were used for deep CT and internal validation of the transferred model, while the left-out test set from the first split was used for external validation of the transferred model.

In the case of the olive data set, the data from instruments 1 and 2 were partitioned according to the schematic presented in Fig. 2. Data from instrument 1 were randomly partitioned into calibration (80 %) and test set (20 %) using the “test_train_split()” function from the sklearn (https://scikit-learn.org/stable/) python library. Afterwards, the calibration set was further partitioned into training (70 %) and validation sets (30 %). Prior to model training, these sets were augmented using the same recipe previously described for the tablet data. The test set from the first split was used exclusively for external validation of the primary model. Data from instrument 2 were partitioned in the same fashion, i.e., fine-tune (80 %) and test set (20 %). Later, the fine-tune set was partitioned into “fine-tune training” (70 %) and “fine-tune validation” sets (30 %), followed by the mentioned data augmentation procedure. The fine-tune training and fine-tune validation set were used for deep CT and internal validation of the transferred model, while the left-out test set from the first split was used for external validation of the transferred model.

2.3. 1-D CNN modelling

To implement the deep learning models, the Python (3.6) language and the open-source deep learning framework TensorFlow/Keras (2.5.0-dev20201204) were used. Keras is a high-level framework that enables a faster and easier implementation of TensorFlow’s capabilities. In this study, we wanted to emphasize the method behind the concept of CT for DL models and for that we opted to use a CNN architecture derived from [18,37] and further explored in [33,38]. Since spectra are represented by 1D vectors, the implemented model is a 1-dimensional convolutional neural network (1D-CNN) composed by 1 conv. layer followed by 4 FC layers (Fig. 3A). Instead of using a feature extraction block composed by a series of conv. and pooling layers like is commonly used in many 2D-CNNs used for image recognition problems, this architecture uses a single conv. layer with 1 filter and a stride of 1. The decision of not implementing pooling layers is motivated by the small size of input

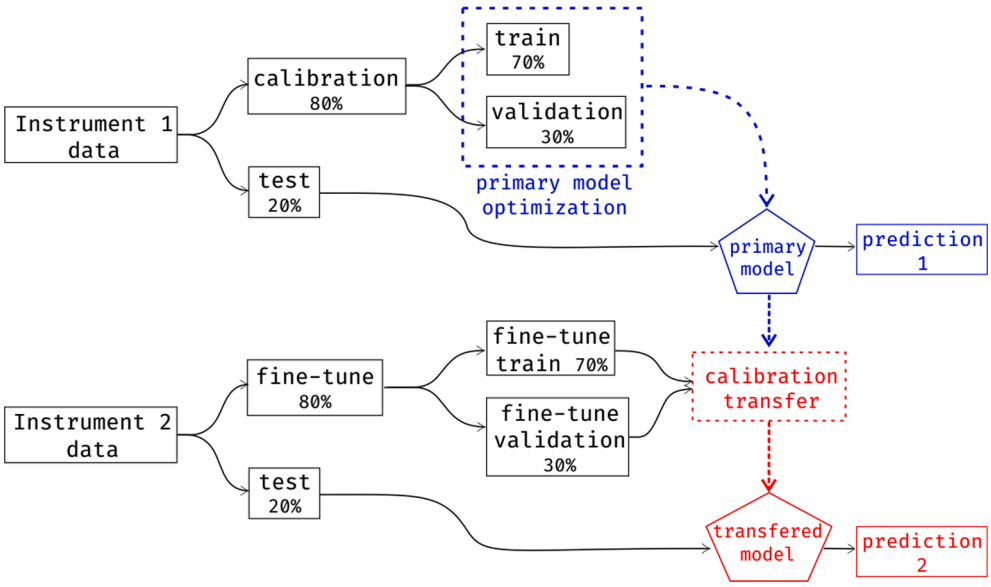

Fig. 2. Data partition scheme for the calibration transfer problem for the olive data set.

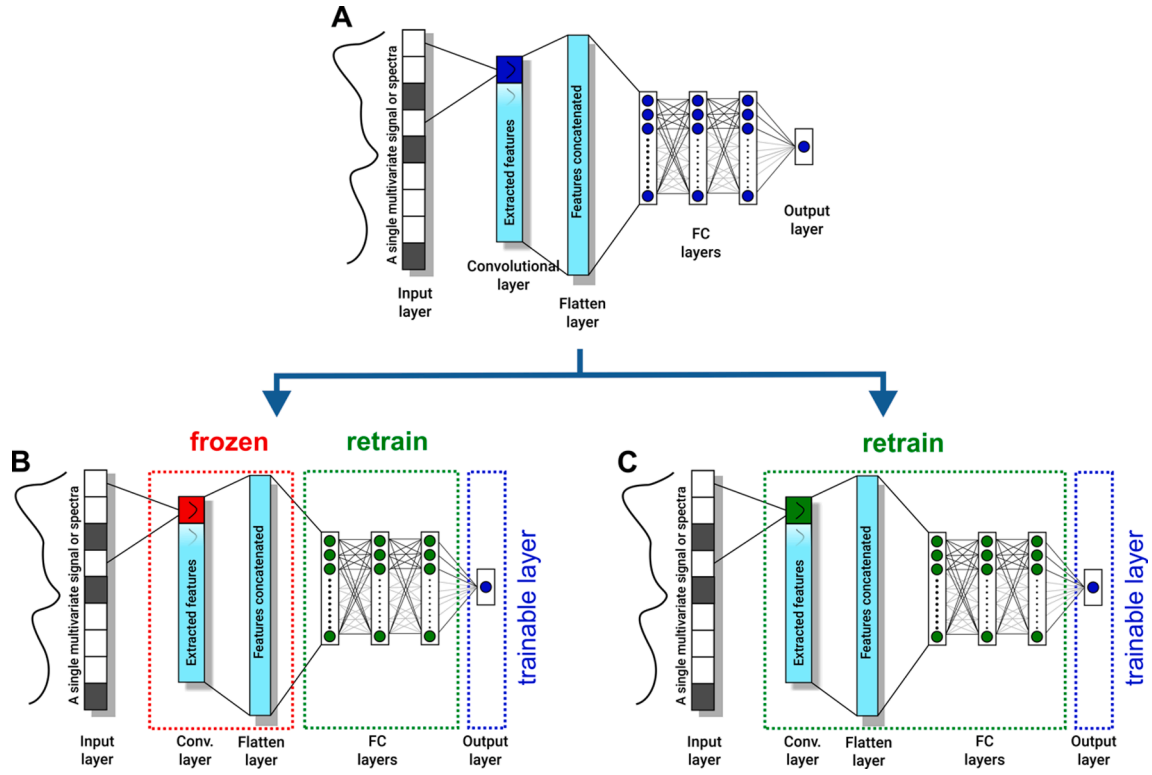

Fig. 3. A summary of the deep CT approach. A) Schematic representation of the CNN architecture used for building the primary model (all trainable layers in blue). B) the primary model is fine-tuned based on some data from new instrument, with the conv. layer kept frozen (red) and retraining the FC block (green). C) The primary model is fine-tuned based on some data from new instrument, by retraining the conv. layer and FC block (green). In all cases, the last layer is always trained from scratch (blue).

vectors and the shallow depth of the 1D-CNN. Polling layers can lead to information loss but are especially useful to perform dimensionality reduction of the output of conv. layers with multiple filters to decrease the overall number of parameters in the model. The 1 filter conv. layer in the model acts as a feature enhancement / pre-processing block. During training, input data is passed on to the conv. layer and the model learns (through back-propagation) the best shape for the filter that performs the feature enhancement. The convolution operation between input data and filter is followed by a non-linear activation function using an exponential linear unit (ELU). The transformed data then flows through the following 3 FC layers with 36, 18 and 12 neurons and ELU activations functions, respectively. In the end, is a single neuron FC layer with a linear activation function to deal with the regression problem. To increase the convergence of the optimization algorithm, we used the "He_normal" procedure to initialize the weights in all the layers. For the sake of reproducibility, the "He_normal" function, and all python and

Tensorflow functions that rely on a random seed number for initialization were instantiated using seed = 42. In this study, the adaptive moment optimizer algorithm (Adam) [39] with an adaptive learning rate was used for the model optimisation. The optimisation algorithm was initially instantiated with a learning rate heuristic [18], LR = 0.01 × (batch size)/256, which was iteratively reduced during the training process (a.k.a. Reduce LR on Plateau). This procedure decreases the convergence time in the initial phase of training and, by reducing progressively the LR step, enables the Adam algorithm to approach minima. The used loss function was the same one defined in [18], i.e., the mean squared error (MSE) added to an L2 penalty on the model weights (layer regularization). This layer regularization helps to prevent model overfitting and is controlled by hyper-parameter β that changes the strength of the L2 penalty. Moreover, to further avoid overfitting, an 'Early Stopping' approach was used. This strategy consists of watching the loss function during training and enforcing a policy that stops the training if, after a certain number of epochs, the validation loss stabilizes or increases. With this procedure in place, the maximum number of training epochs allowed (700 in this study) was never achieved.

To reach optimal DL models, three different hyperparameters were tuned (width of the convolution filter (kernel width), the strength of the L2 regularization (β) and the training batch size) using a grid-search strategy [33]. A summary of the search space is shown in Table 1.

2.4. Transfer learning for calibration transfer – Deep CT

The CNN model used in this study has three main parts, i.e., feature enhancement/extraction using a conv. layer, a mapping block with FC layers, and the final output layer that holds the prediction for the target variable. The TL method widely used in the DL computer vision domain consists in adapting the weights of certain layers of a pre-trained NN model, by showing the NN a new subset of images of the target domain. For example, a NN can be trained to generally identify faces in pictures, and, through TL, it can be fine-tuned to find smiley faces in a new subset of pictures [40,41]. Depending on the NN architecture and final purpose, TL can be accomplished by re-training certain blocks of the NN while keeping other blocks with their pre-trained weights (frozen layers). This can be seen as implementing different TL strategies.

Following what was mentioned in the introduction section, in the case of the instrument transfer presented in this study, freezing the conv. layers might not have equivalent results as their 2D-CNN computer vision aimed counterparts. This is because the instrument differences in many cases are local and may require retraining of the weights of the conv. layer. Furthermore, the weights in the FC layers should not be completely replaced as they have already learned some useful mapping relationships between the features extracted by the conv. layers and the response variables. Retraining a layer for a few epochs (on a smaller data subset) can slightly modify the existing weights in the layers/neurons, or in broader terms, it just adapts the already learned knowledge encoded in the neuron's weights to the new data.

In this work two TL strategies were implemented and compared. For both cases the block of FC layers was initialized with the weights of the primary model (based on instrument 1) and retrained using the fine-tune data from instrument 2, while the last FC layer was allowed to train from scratch (i.e., initialized with random weights). In one case the conv. layer was kept frozen (Fig. 3B), i.e., the weights of the primary model where kept, while, in the other case, the conv. layer could be re-trained (like the FC layers), i.e., initialized with the weights from the primary model (Fig. 3C). The two separate cases are highlighted in Fig. 3.

Following the scheme presented in Fig. 1, in both cases, the 1D-CNN model optimised and trained for the primary instrument was used as the starting point for the deep CT. The fine-tune data set, as explained in the 'data augmentation' section was used for adapting the primary model to instrument 2. Finally, the performance of the transferred model was tested on the external test set as explained in the 'data augmentation' section. The performances of the models were judged based on the root mean squared error (RMSE).

Table 1
Intervals hyperparameters (HP) optimisation.

Name	Interval / step
Conv. filter size	[5 –30]
L2 regularization β	[0.001, 0.003, 0.008, 0.01, 0.015, 0.02, 0.03]
Batch size	[32, 64, 128, 256, 512]

3. Results and discussion

3.1. Spectra from different cases

The mean spectra of the tablets and olive from instruments 1 and 2 are shown in Fig. 4A and 4B, respectively. For the tablet data set, the differences in the mean responses of the instruments can be found at several locations such as ~ 700 nm, ~ 800 nm, ~ 1300 nm and ~ 1650 nm. Such localised difference can be related to the intrinsic differences in the instrument detectors (Fig. 4A). For the olive data set, the difference is global and can be noticed along the complete spectral range (Fig. 4B). Such global differences can arise from a range of factors such as differences in light source, instrument detector and measurement conditions. The main point to note is that if the model made on data from one instrument is used on data of a new instrument, it will perform badly due to the existing differences between instruments. Hence, a model transfer compensating the instrument difference is needed.

3.2. Transfer between bench top instruments – Tablet data set

At first, the primary 1D-CNN model based on instrument 1 was developed to predict (active pharmaceutical ingredients) API assay in tablets. The optimization grid-search found that models with β = 0.02 showed the lowest RMSE for both training and validation set. Furthermore, for β = 0.02, a kernel size of 20 and batch size of 64 were chosen as reflecting minima on both training (Fig. 5A) and validation set (Fig. 5B).

The 1D-CNN model based on the optimal parameters (Fig. 5) showed a RMSE of 3.513 mg on the test set for the instrument 1 (Fig. 6A). However, using the model made on instrument 1 on the test set of instrument 2 showed a RMSE of 13.059 mg (Fig. 6B). Such a drastic increase in the RMSE was due to the unmodeled differences between instrument 1 and 2. After that, the model based on instrument 1 was adapted to be used on instrument 2 using deep CT. The results from the transferred model are shown in the 2nd row of Fig. 6. It can be noted that both the model transfer approaches (Fig. 3) were able to reduce the RMSE from 13.059 mg (on data from instrument 2) to the same level as the RMSE of the primary instrument i.e., ~ 3.513 mg. Moreover, the performance of model transferred while allowing the tuning of conv. layer improved (RMSE = 3.258 mg) when compared to the model transferred while keeping the conv. layer frozen (RMSE = 3.608 mg).

3.3. Transfer between handheld instruments – Olive data set

Again, like in the previous experiment, the first step was the development of a primary 1D-CNN model using on data from instrument 1 to predict DM in olive fruit. The grid-search optimization procedure identified β = 0.015 as showing the lowest RMSE for both training and validation set. For this L2 regularization value, the RMSE minimum in the kernel size vs batch size space is found for kernel size = 30 and batch size = 64 on both training (Fig. 7A) and validation set (Fig. 7B).

The 1D-CNN model based on the optimal parameters (Fig. 7) showed a RMSE of 0.629 % on the test set for the instrument 1 (Fig. 8A). However, using the model made on instrument 1 on the test set of instrument 2 showed an increased RMSE of 0.858 % (Fig. 8B). The relative increase in RMSE was lower compared to the tablet case. A reason for this could be the absence of local differences (i.e., in specific bands) in

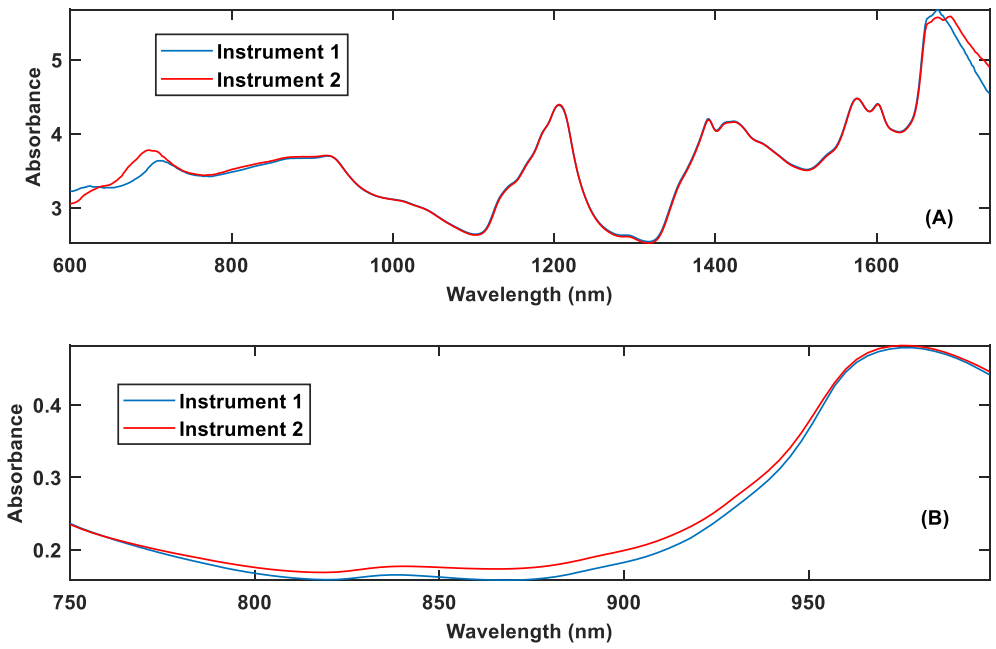

Fig. 4. The mean spectrum for tablets from two different lab-based instruments (A), and the mean spectrum of olive from different hand-held instruments (B).

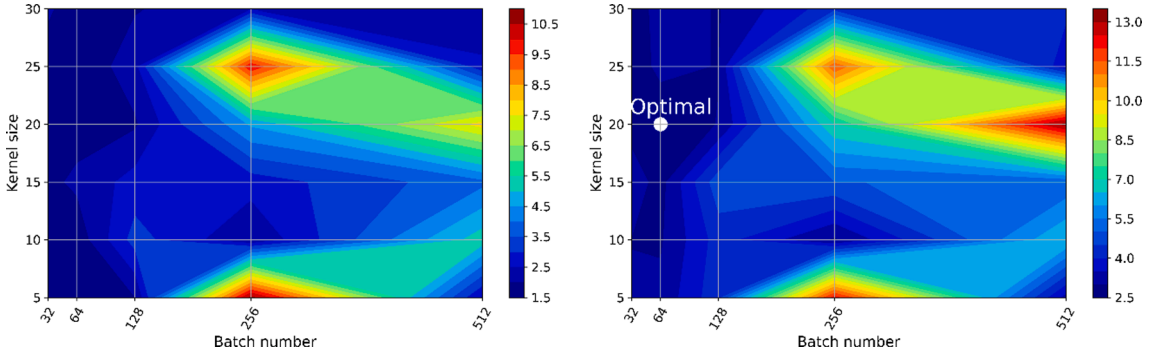

Fig. 5. RMSE hyperparameter space for model train (A) and validation (B) errors for L2 regularise (β = 0.02). The optimal point showing lowest error is highlighted as optimal in (B).

the instruments for the olive case (more on Section 4.4). The model based on instrument 1 was then transferred to be used on instrument 2 using the deep CT. The results from the transferred model are shown in the 2nd row of Fig. 7. Both model transfer approaches were able to reduce the RMSE from 0.858 % to like the RMSE of the level of instrument 1 i.e., ~0.629 %. The performance of the model transferred while allowing the tuning of conv. layer was marginally better (RMSE = 0.658 %) compared to the model transferred while keeping the conv. layer frozen (RMSE = 0.665 %).

3.4. Difference learned by models

In this study, two types of deep CT approaches were tested. The first approach allowed the conv. layer to be retrained based on the data from the new instrument, while the second approach kept conv. layer frozen. In the previous section, the retraining of the conv. layer showed better model performance (for both olive and tablet data sets) when compared to keeping the conv. layer frozen (Table 2). To understand how the retraining of conv. layers allowed better model performance, the mean activation of the conv. layer in the primary model made on instrument 1 and the transferred models are shown in Fig. 9 (tablet data) and Fig. 10 (olive data). For tablet data, it can be noted that the transferred model (retraining the conv. layer) was able to adapt to the local differences (600–700 nm, 1150–1200 nm and 1300–1400 nm bands) present in the instruments (as shown in Fig. 4A). For the olive case (Fig. 10), the transferred model (retraining the conv. layers) was able to adapt to the global differences (baseline shift effect) present in the instruments (as shown in Fig. 4B). Hence, the retraining of the conv. layers in these cases seems to be necessary to correctly model the instrument differences. Another evidence for supporting these explanations comes from the size of the learned kernels for both olive and tablet data. The kernel width for olive data is broader than that for the tablet data. Convolution operations with larger kernels usually translate in more global data transformations such as baseline corrections and overall amplitude shifts.

3.5. Posterior analysis of the effect of sample size on the performance of deep calibration transfer

So far, it was shown that the deep CT approach was able to regain the predictive performance of DL models when transferred from one instrument to another. A key step in this technique is that it requires some new samples measured on the new instrument. In the cases presented, the available fine-tune set consist of 194 samples for the tablet data set and 351 samples for the olive data set. In many cases, it can be assumed that the user may not have access to many samples and will like to use as low samples as possible to reduce the experiment burden to save time

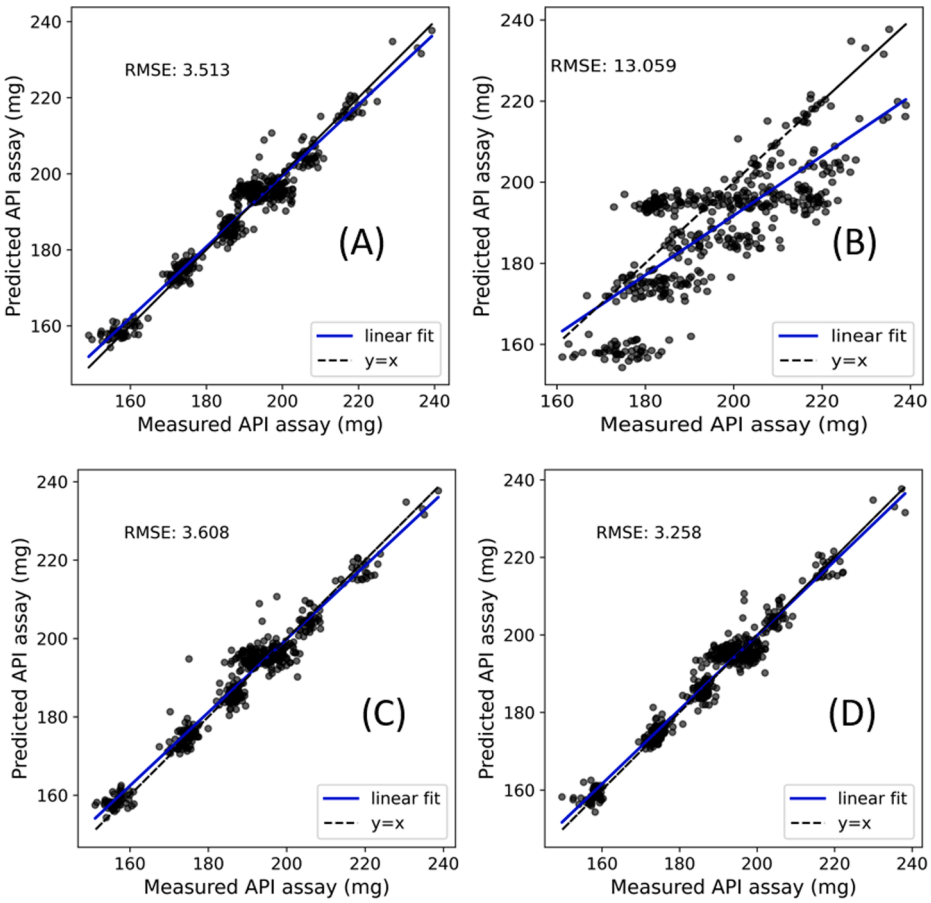

Fig. 6. 1D CNN model made on instrument 1 and tested on test data from instrument 1 (A) and instrument 2 (B). Transferred model with frozen conv. layer tested on test data from instrument 2 (C) and transferred model with conv. layer tuning tested on test data from instrument 2 (D).

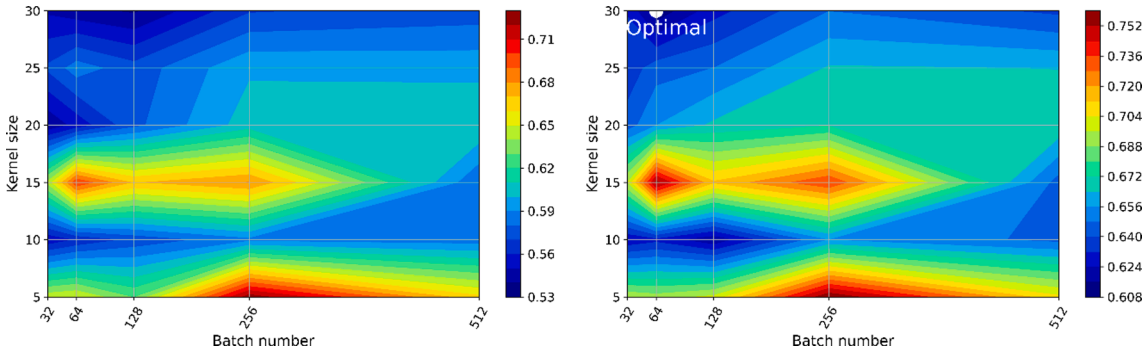

Fig. 7. A summary of model training (A) and validation (B) errors for L2 regularise (β = 0.015). The optimal point showing lowest error is highlighted as optimal in (B).

and costs. Hence, to identify how the sample size of fine-tune set affects the performance of deep CT, a posterior analysis on the tablet data set was performed and presented in Fig. 11. In this figure, it can be noted that initially, with a small samples size of 10 – 40, the model performance was poor. However, with increasing samples size the RMSEP was decreased. Furthermore, a samples' size of 76 was able to attain the RMSEP = 3.42 mg, lower than the RMSEP = 3.51 mg on the primary instrument as reported in Fig. 6A. Hence, the posterior analysis concludes that instead of using the 181 samples available in the fine-tune set, the analysis with 76 samples is sufficient to attain the RMSEP like the primary instrument.

A key point to note is that although in this study, for demonstration purposes, the primary DL models transferred to the new instruments were based on a limited amount of data, the practical scenarios where DL models are usually deployed involve large data sets (which the users have acquired for a long time or in larger experiments). In the case of calibration transfer, it can be assumed that a user may have acquired lots of data on a particular instrument for many years and many experiments and then he/she developed a global DL model of such valuable data. If by chance the old instrument got damaged, then the user cannot repeat all the years of experiments on which they acquired such valuable data. Hence, in that case, the presented method i.e., deep CT comes into play with which a small set of data can allow the user to fine-tune their old model and use it on the new instrument.

To properly train a neural network (NN), two main steps are required: the first is to find the NN architecture and its hyperparameters

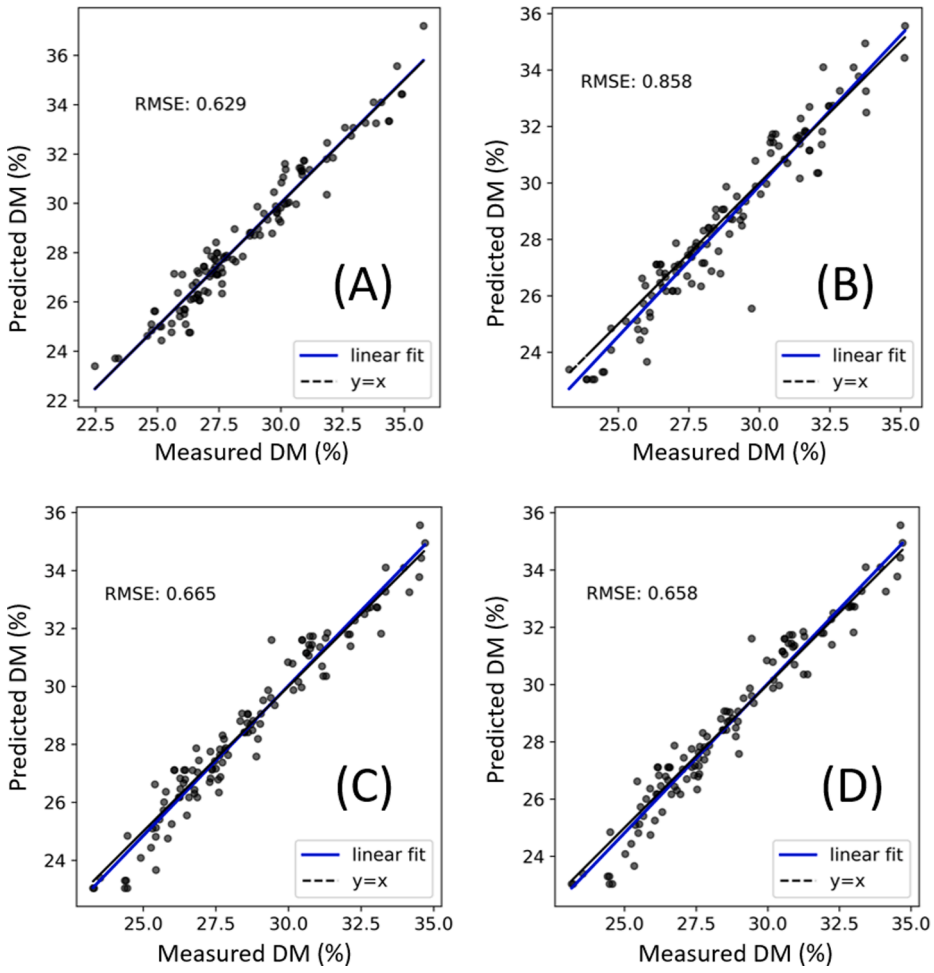

Fig. 8. 1D CNN model made on instrument 1 and tested on test data from instrument 1 (A) and instrument 2 (B). Transferred model without conv. layer tuning tested on test data from instrument 2 (C) and transferred model with conv. layer tuning tested on test data from instrument 2 (D).

Table 2
A summary of model performance before and after model training.

	Root mean squared error of prediction (RMSEP)			
	Primary instrument	Primary model applied to new instrument	Deep CT without conv. layer training	Deep CT with conv. layer training
Transfer between lab instruments	3.513	13.059	3.608	3.258
Transfer between portable handheld instruments	0.629	0.858	0.665	0.658

that allow the model to achieve good learning capabilities but after that, on a more fundamental level, the "learning process" itself is done during model training by finding the weights and biases for each unit/neuron (on every layer of the model) that allows mapping the input features into the target variable. It can be understood as the distribution or pattern of the learned weights as the final model. When TL is applied with the proposed approach in this study, the learning process is done by fine-tuning these pre-learned weights by retraining the model on a small number of new data samples. Based on the findings of this study, a small number of data samples were enough to expand the "information/knowledge" of the NN. A key point to note is that the TL approach requiring fine-tuning of the existing DL models should not be confused with the retraining of the model from scratch. As retraining from scratch means initializing all units' weights as random numbers and the smaller/newly available data set is not usually enough to train all weights properly.

In this study, two cases of CT were demonstrated covering both the lab-based instrument and the portable hand-held spectrometers. The two demonstrated cases were carefully selected as they cover both the global as well as local differences in the instrument responses. However, one can also assume that in some special cases, the transfer of model may also be required from a lab-based instrument to a portable instrument. In principle, the proposed deep CT approach can also be implemented as it can handle both the global and the local differences. However, its performance still must be investigated further in order to ensure that deep CT can handle corrections for both types of differences simultaneously. Prior to the implementation of the deep CT, pre-processing steps such as spectral registration and interpolation may be required to match the spectral ranges from a lab-based and a portable instrument.

4. Conclusions

This study showed that, like chemometrics classic models, DL models also have generalisation problems when directly used on a new instrument. Therefore, the DL models also require a model transfer procedure to adjust to the new instrument on which they are intended to be deployed. To deal with the transfer of DL models between instruments,

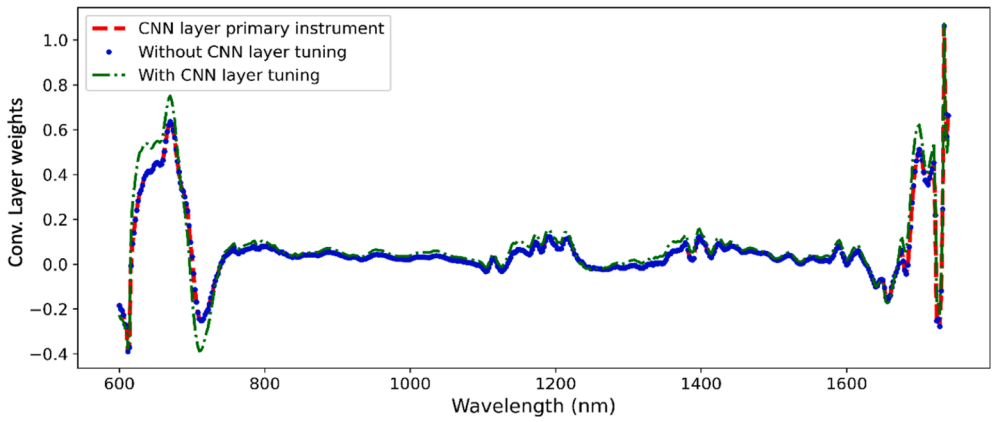

Fig. 9. Mean conv. layer activations for API assay prediction in tablets. Activation of old model (red dashed), activation of transferred model without conv. layer tuning (blue dots), and activation of transferred model with conv. layer tuning (green dashed dotted).

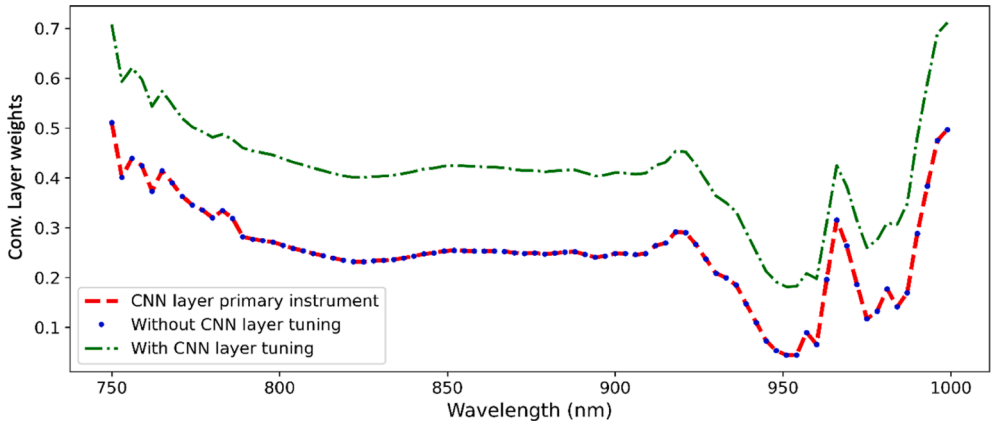

Fig. 10. Mean of conv. layer activations for DM prediction in olives. Activation of old model (red dashed), activation of transferred model without conv. layer tuning (blue dots), and activation of transferred model with conv. layer tuning (green dashed dotted).

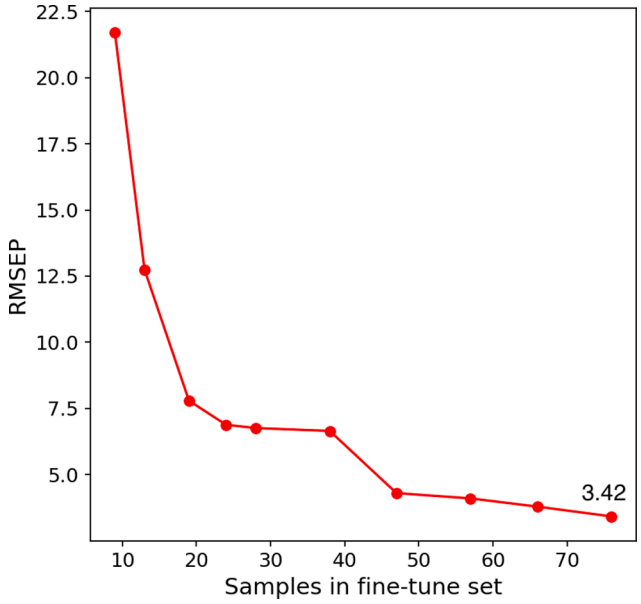

Fig. 11. Effect of samples size of fine tune set on the deep CT model. A samples size of 76 was sufficient to attain the RMSEP like the primary instrument.

this study proposes deep CT. The deep CT method is based on the concept of DL transfer learning for model fine-tuning where the primary DL model was retrained with some new data measured on a new instrument. Further, it was showed that the deep CT can deal with both local and global differences between the instruments by allowing the retrain the conv. layer in the CNN. For both the lab-based and hand-held spectrometers, the transferred model was able to regain the predictive performance. Deep CT has the potential to allow users of spectroscopy to keep reusing their DL models on different instruments. For that, the first steps presented here should be confirmed and validated through its application to more data sets as they become available. Furthermore, the deep CT is a standard-free approach, hence can be used in cases when the primary instrument is unavailable. One of the limitations of the present study is that, due to the unavailability of a multi-instruments large data sets, the experiment was performed on augmented data and the full capability of Deep CT was not demonstrated. The impact of the data augmentation procedure, for DL models in general, needs to be better characterized in the case of spectral data. However, with such a satisfactory performance on these small data sets, it is expected that in the availability of large data sets and more complex/deeper models, the deep CT performance should improve given the known performance gains DL models shows with more data. The main disadvantage of the approach is that it requires some new measurements on the new instrument both the spectra and reference property of interest to perform the model transfer. Future investigations should focus on research directions that allow adapting the deep spectra models to a new instrument without the need of new reference property measurements.

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.

References

[1] A. Biancolillo, F. Marini, Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis, Front. Chem. 6 (2018) 576.
[2] L.E. Agelet, C.R. Hurburgh, A Tutorial on Near Infrared Spectroscopy and Its Calibration, Crit. Rev. Anal. Chem. 40 (2010) 246–260.
[3] P. Mishra, F. Marini, B. Brouwer, J.M. Roger, A. Biancolillo, E. Woltering, E.H.-v. Echtelt, Sequential fusion of information from two portable spectrometers for improved prediction of moisture and soluble solids content in pear fruit, Talanta 223 (2021) 121733.
[4] P. Mishra, E. Woltering, B. Brouwer, E. Hogeveen-van Echtelt, Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach, Postharvest Biol. Technol. 171 (2021) 111348.
[5] C. Pasquini, Near infrared spectroscopy: A mature analytical technique with new perspectives – A review, Anal. Chim. Acta 1026 (2018) 8–36.
[6] R. Nikzad-Langerodi, F. Sobieczky, Graph-based calibration transfer, arXiv preprint arXiv:2006.00089, (2020).
[7] R. Nikzad-Langerodi, W. Zellinger, S. Saminger-Platz, B.A. Moser, Domain adaptation for regression under Beer–Lambert's law, Knowledge-Based Systems, 210 (2020) 106447.
[8] Puneet Mishra, Ramin Nikzad-Langerodi, Partial least square regression versus domain invariant partial least square regression with application to near-infrared spectroscopy of fresh fruit, Infrared Phys. Technol. 111 (2020) 103547, https://doi.org/10.1016/j.infrared.2020.103547.
[9] P. Mishra, J.M. Roger, D.N. Rutledge, E. Woltering, Two standard-free approaches to correct for external influences on near-infrared spectra to make models widely applicable, Postharvest Biol. Technol. 170 (2020) 111326.
[10] J.J. Workman, A Review of Calibration Transfer Practices and Instrument Differences in Spectroscopy, Appl. Spectrosc. 72 (2017) 340–365.
[11] B. Malli, A. Birlutiu, T. Natschläger, Standard-free calibration transfer - An evaluation of different techniques, Chemo. & Intel. Lab. Syst. 161 (2017) 49–60.
[12] A. Folch-Fortuny, R. Vitale, O.E. de Noord, A. Ferrer, Calibration transfer between NIR spectrometers: New proposals and a comparative study, J. Chemom. 31 (2017) e2874.
[13] R.N. Feudale, N.A. Woody, H. Tan, A.J. Myles, S.D. Brown, J. Ferré, Transfer of multivariate calibration models: a review, Chemo. & Intel. Lab. Syst. 64 (2) (2002) 181–192.
[14] P. Mishra, R. Nikzad-Langerodi, F. Marini, J.M. Roger, A. Biancolillo, D.N. Rutledge, S. Lohumi, Are standard sample measurements still needed to transfer multivariate calibration models between near-infrared spectrometers? The answer is not always, TrAC Trends in Analytical Chemistry, (2021) 116331.
[15] L. Sun, C. Hsiung, V. Smith, Investigation of Direct Model Transferability Using Miniature Near-Infrared Spectrometers, Molecules 24 (10) (2019) 1997, https://doi.org/10.3390/molecules24101997.
[16] Y. Wang, D.J. Veltkamp, B.R. Kowalski, Multivariate instrument standardization, Anal. Chem. 63 (1991) 2750–2756.
[17] Y. Wang, B.R. Kowalski, Calibration Transfer and Measurement Stability of Near-Infrared Spectrometers, Appl. Spectrosc. 46 (1992) 764–771.
[18] Chenhao Cui, Tom Fearn, Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration, Chemo. & Intel. Lab. Syst. 182 (2018) 9–20.
[19] Z. Xin, S. Jun, T. Yan, C. Quansheng, W. Xiaohong, H. Yingying, A deep learning based regression method on hyperspectral data for rapid prediction of cadmium residue in lettuce leaves, Chemo. & Intel. Lab. Syst. 200 (2020) 103996.
[20] X.J. Yu, H.D. Lu, D. Wu, Development of deep learning method for predicting firmness and soluble solid content of postharvest Korla fragrant pear using Vis/NIR hyperspectral reflectance imaging, Postharvest Biol. Technol. 141 (2018) 39–49.
[21] X. Yu, H. Lu, Q. Liu, Deep-learning-based regression model and hyperspectral imaging for rapid detection of nitrogen concentration in oilseed rape (Brassica napus L.) leaf, Chemo. & Intel. Lab. Syst. 172 (2018) 188–193.
[22] S.J. Pan, Q. Yang, A Survey on Transfer Learning, IEEE Trans. Knowl. Data Eng. 22 (2010) 1345–1359.
[23] K. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning, J. Big Data 3 (2016) 9.
[24] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He, A comprehensive survey on transfer learning, Proc. IEEE 109 (2020) 43–76.
[25] J. Padarian, B. Minasny, A.B. McBratney, Transfer learning to localise a continental soil vis-NIR calibration model, Geoderma 340 (2019) 279–288.
[26] N. Zhu, X. Ji, J. Tan, Y. Jiang, Y. Guo, Prediction of dissolved oxygen concentration in aquatic systems based on transfer learning, Comput. Electron. Agric. 180 (2021) 105888.
[27] W. Xu, G. Yu, A. Zare, B. Zurweller, D.L. Rowland, J. Reyes-Cabrera, F.B. Fritschi, R. Matamala, T.E. Juenger, Overcoming small minirhizotron datasets using transfer learning, Comput. Electron. Agric. 175 (2020) 105466.
[28] B. Espejo-Garcia, N. Mylonas, L. Athanasakos, S. Fountas, I. Vasilakoglou, Towards weeds identification assistance through transfer learning, Comput. Electron. Agric. 171 (2020) 105306.
[29] J. Chen, J. Chen, D. Zhang, Y. Sun, Y.A. Nanehkaran, Using deep transfer learning for image-based plant disease identification, Comput. Electron. Agric. 173 (2020) 105393.
[30] W.M. Kouw, M. Loog, An introduction to domain adaptation and transfer learning, arXiv preprint arXiv:1812.11806, (2018).
[31] A. Khan, A. Sohail, U. Zahoora, A.S. Qureshi, A survey of the recent architectures of deep convolutional neural networks, Artif. Intell. Rev. 53 (2020) 5455–5516.
[32] L. Liu, M. Ji, M. Buchroithner, Transfer Learning for Soil Spectroscopy Based on Convolutional Neural Networks and Its Application in Soil Clay Content Mapping Using Hyperspectral Imagery, Sensors (Basel, Switzerland) 18 (2018) 3169.
[33] P. Mishra, D. Passos, Realizing transfer learning for updating deep learning models of spectral data to be used in a new scenario, Chemo. & Intel. Lab. Syst. (2021) 104283.
[34] X. Sun, P. Subedi, R. Walker, K.B. Walsh, NIRS prediction of dry matter content of single olive fruit with consideration of variable sorting for normalisation pre-treatment, Postharvest Biol. Technol. 163 (2020) 111140.
[35] W. Du, Z.-P. Chen, L.-J. Zhong, S.-X. Wang, R.-Q. Yu, A. Nordon, D. Littlejohn, M. Holden, Maintaining the predictive abilities of multivariate calibration models by spectral space transformation, Anal. Chim. Acta 690 (2011) 64–70.
[36] E.J. Bjerrum, M. Glahder, T. Skov, Data augmentation of spectral data for convolutional neural network (CNN) based deep chemometrics, arXiv preprint arXiv:1710.01927, (2017).
[37] J. Acquarelli, T. van Laarhoven, J. Gerretzen, T.N. Tran, L.M.C. Buydens, E. Marchiori, Convolutional neural networks for vibrational spectroscopic data analysis, Anal. Chim. Acta 954 (2017) 22–31.
[38] P. Mishra, D. Passos, A synergistic use of chemometrics and deep learning improved the predictive performance of near-infrared spectroscopy models for dry matter prediction in mango fruit, Chemo. & Intel. Lab. Syst. 212 (2021) 104287, https://doi.org/10.1016/j.chemolab.2021.104287.
[39] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, (2014).
[40] J. Li, S. Huang, X. Zhang, X. Fu, C.-C. Chang, Z. Tang, Z. Luo, Facial Expression Recognition by Transfer Learning for Small Datasets, Springer International Publishing, Cham, 2020, pp. 756–770.
[41] M.A.H. Akhand, S. Roy, N. Siddique, M.A. Kamal, T. Shimamura, Facial Emotion Recognition Using Transfer Learning in the Deep CNN, Electronics 10 (9) (2021) 1036, https://doi.org/10.3390/electronics10091036.