Chemometrics and Intelligent Laboratory Systems 212 (2021) 104283

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Realizing transfer learning for updating deep learning models of spectral data to be used in new scenarios

Puneet Mishra [a,*], Dário Passos [b]

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

ARTICLE INFO

Keywords:
Transfer Learning
Generalizability
Spectroscopy
Process monitoring

ABSTRACT

This study presents the concept of transfer learning (TL) to the chemometrics community for updating DL models related to spectral data, particularly when a pre-trained DL model needs to be used in a scenario having unseen variability. This is the typical situation where classical chemometrics models require some form of re-calibration or update. In TL, the network architecture and weights from the pre-trained DL model are complemented by adding extra fully connected (FC) layers when dealing with the new data. Such extra FC layers are expected to learn the variability of the new scenario and adjust the output of the main architecture. Furthermore, three approaches of TL were compared, first where the weights from the initial model were left untrained and the only the newly added FC layers could be retrained. The second was when the weights from the initial model could be retrained alongside the new FC layers. The third was when the weights from the initial model could be re-trained with no extra FC layers added. The TL was shown using two real cases related to near-infrared spectroscopy i.e., mango fruit analysis and melamine production monitoring. In the case of mango, the model needs to be updated to cover a new seasonal variability for dry matter prediction, while, for the melamine case, the model needs to be updated for the change in the recipe of the production material. The results showed that the proposed TL approaches successfully updated the DL models to new scenarios for both the mango and melamine cases presented. The TL performed better when the weights from the old model were retrained. Furthermore, TL outperformed three recent benchmark approaches to model updating. TL has the potential to make DL models widely useable, sharable, and scalable.

1. Introduction

Multivariate predictive modelling is widely performed in the chemometrics domain to complement analytical sensing technologies [1]. Several techniques are available, but the most used is the partial least-squares (PLS) regression [2,3] and its variants [4]. PLS is popular because it allows dealing with high multi-collinearity in the data as well as many variables in the input signals by modelling latent spaces that are correlated to the property of interest [2,5]. However, recently, neural networks (NN) in the framework of deep learning (DL) are appearing as a potential tool for performing multivariate analysis [6–10]. Current DL methodologies for multivariate predictive modelling can be divided into two main categories. The first is a supervised approach and involves joint feature extraction and modelling with convolutional neural networks (CNN), that combine convolution layers (CONV) with fully connected (FC) layers [6,10]. The second approach involves an unsupervised extraction of features using techniques such as autoencoders, and later, using those features with another NN or on classical regression algorithms such as support vector machine (SVM) to relate to the property of interest. DL approaches have already outperformed the PLS analysis in several demonstrated cases [7,8,11]. However, unlike PLS analysis, current DL methods can only be implemented in the availability of large data sets (thousands of sample points and corresponding reference measurements) [6,10]. Thanks to the advancement in sensing technologies, such large data sets are getting increasingly popular and are being openly shared by the scientific community. Currently, two domains where these data sets are available are the portable consumer near-infrared (NIR) spectroscopy [12–15] and industrial process monitoring [16].

Chemometric calibrations are highly specific and often require updates to accommodate changes in the data originated by any change in physical, chemical, or surrounding environmental conditions [16–19]. In

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

https://doi.org/10.1016/j.chemolab.2021.104283
Received 29 December 2020; Received in revised form 8 February 2021; Accepted 28 February 2021
Available online 5 March 2021

other words, chemometric methods only work well in situations for which the model already incorporates the variability seen in the calibration data. Furthermore, the performance of standard machine learning and chemometric models tend to saturate (create a plateau) as the size of data increases. On the other hand, DL models keep improving their performance with the increasing number of samples leading to a better prediction [6,10]. However, in the case of new unseen variability in the data, just like standard chemometric methods, the DL approaches do not precisely estimate the property of interest. In that case, the DL model must be updated or adapted to capture the new variability. In the domain of DL for computer vision, the problem of model update or adaption is widely explored and understood [20]. Several global models such as AlexNet [21], VGGNet [22], ResNet [23] and inception net [24], trained using millions of images, are often updated or adapted for specific applications [25–30]. For example, a user interested in developing a computer vision model for his specific application does not require millions of images to train the model from scratch, but just requires a new subset of images and pre-trained model weights from any global model [28]. Afterwards, the user can implement fine-tuning of model weights (of the global model) based on the new image set, and some fully connected (FC) layers to adapt model weights [26]. This task of fine-tuning a pre-trained model is commonly referred to as transfer learning (TL) in the DL community. A formal description of the concept of TL in the context of machine learning models can be found in Ref. [41] and, in recent reviews [42,43], highlighted the many areas where this concept is being applied. Like computer vision, TL can also be performed for updating the DL models related to spectral data. TL for spectral data can be achieved by using the default network architecture with the weights from a pre-trained model, adding extra FC layers and some measurements from a new scenario to perform the model fine-tuning. Even though the application of TL in computer vision problems has become a standard process, to the best of our chemometric literature search, TL methods for spectral data modelling have not been developed yet, and therefore, this study is the first to show its capability on real spectroscopy data sets.

The study aims to present three TL approaches for updating DL models related to spectral data processing. The TL approaches were based on the weight transfer and fine-tuning, where for a new scenario, the model architecture and weights from the pre-trained model are complemented by adding extra FC layers. Such extra FC layers are expected to learn the variability of the data originated in the new scenario as well as adjust the output of the main architecture accordingly to it. Further, three approaches of TL were compared; in the first, the model weights from the old architecture were left untrained and the new learning was performed by the extra FC layers; in the second, the model weights from the old architecture could be trained alongside the extra FC layers; in the third, the model weights from the old architecture were trained without the addition of extra FC layers. TL was showed using two real cases related to near-infrared spectroscopy i.e., mango fruit analysis and melamine production monitoring. In the case of mango, the model needs to be updated to cover the variability in dry matter prediction induced by a new harvest season, while, for the melamine case, the model needs to be updated to account for changes in the recipe of the melamine production. As a baseline, the results were compared with three different model updating approaches. Out of the three approaches, two were chemometric, namely recalibration of PLS with some data from the new scenario as performed in Ref. [31] and the recently proposed semi-supervised parameter-free calibration enhancement framework (PFCE) [32]. The third method is based on DL models weight sharing approach for non-translation data [33].

2. Materials and method

2.1. Data sets

2.1.1. Mango dry matter prediction data set

The mango data set used in this study includes 11,691 NIR spectra (684–990 nm) and DM measurements performed on 4675 mango fruit across 4 harvest seasons 2015, 2016, 2017 and 2018 [14,15]. A portable F750 Produce Quality Meter (Felix Instruments, Camas, USA) was used for the non-destructive NIR measurements. DM was measured with oven drying (UltraFD1000, Ezidri, Beverley, Australia). Out of 11,691 spectra, 10,243 spectra corresponding to harvest seasons 2015, 2016 and 2017 were used as a train set for developing the primary model, while the remaining 1448 spectra from season 2018 were used for fine-tuning and updating the primary model. The spectra from the season 2018 were further divided into fine-tuning (60%) and independent testing (40%) sets using Kennard Stone algorithm [34]. The abnormal samples were removed from the analysis using hoteling T^2 and Q statistics. The original data set can be accessed at: https://data.mendeley.com/datasets/46htwnp833/1.

2.1.2. Melamine production data set

The melamine dataset comprises NIR absorbance spectra and turbidity point (expressed in °C) readings of melamine formaldehyde recorded during condensation (i.e., polymerization) process at Metadynea GmbH (Krems, Austria) [35,36]. The spectra were recorded in-line through a fiber optical probe while the turbidity point was measured offline i.e., the temperature at which the condensate starts to get turbid. The turbidity point is a measure of polymer length and amount of cross-linking (i.e., the degree of polymerization) and is an important quantity that defines the physical and chemical properties of the final resin. The dataset includes recordings on several batches from 4 different recipes that use different raw material ratios and/or additives. In the original data, 4 different recipes were mentioned as 'R562','R568′,'R861′ and 'R862' and has 3032, 733, 3890 and 462 spectral measurements and reference turbidity points, respectively. To ease the DL task the data from 3 recipes ('R568′,'R861′ and 'R862') were combined as a train set, with a total of 5085 spectra and reference turbidity points, for optimizing the primary model making. The data related to recipe 'R562' was used for model fine-tuning (60%) and independent testing (40%) and was partitioned using the KS algorithm [34]. Like in the previous data set, abnormal samples were removed from the analysis using hoteling T^2 and Q statistics. The original data set can be accessed at: https://github.com/RNL1/Melamine-Dataset.

2.2. Deep learning

For the initial model development phase, the train data set was further subdivided into calibration (75%) and tuning (25%) sets using the train-test split function of SciKit learn (https://scikit-learn.org/stable/). In this study, a 1-dimensional convolutional neural network (1D-CNN) with a layer structure like the one suggested by Ref. [6] was used. The model was implemented using the Python (3.6) language and TensorFlow/Keras (2.5.0-dev20201204) running on a workstation equipped with a NVidia GPU (GeForce RTX 2080 Ti), an Intel® Core™ i7-4770k @3.5 GHz and 64 GB RAM, running Microsoft Windows 10 OS.

The initial CNN is constituted by 1 input layer, 1 convolutional layer with 1 kernel (a.k.a filter) and stride = 1, 4 fully connected (FC) layers with 36, 18, and 12 neurons, respectively, and a single output value. An illustration of the network is shown in Fig. 1A. Since this CNN was implemented for regression, a linear activation function was used in the output layer while exponential linear unit (ELU) activation functions were used in all the others. The weights in all layers of the initial CNN were initialized using the 'He_normal' initialization procedure [37]. All models were trained using an adaptive moment optimizer algorithm (Adam) [38] with an initial learning rate (LR) given by LR = 0.01 × (batch size)/256. As [6] suggests, the use of this empirical rule allows a good performance balance between learning rate (LR) and batch size. Lower LR values tend to increase the time of convergence of the Adam algorithm towards minima, and larger values usually decrease the algorithm's capability to get closer to a minimum value. In this study, the initial LR was estimated using the above rule but was automatically

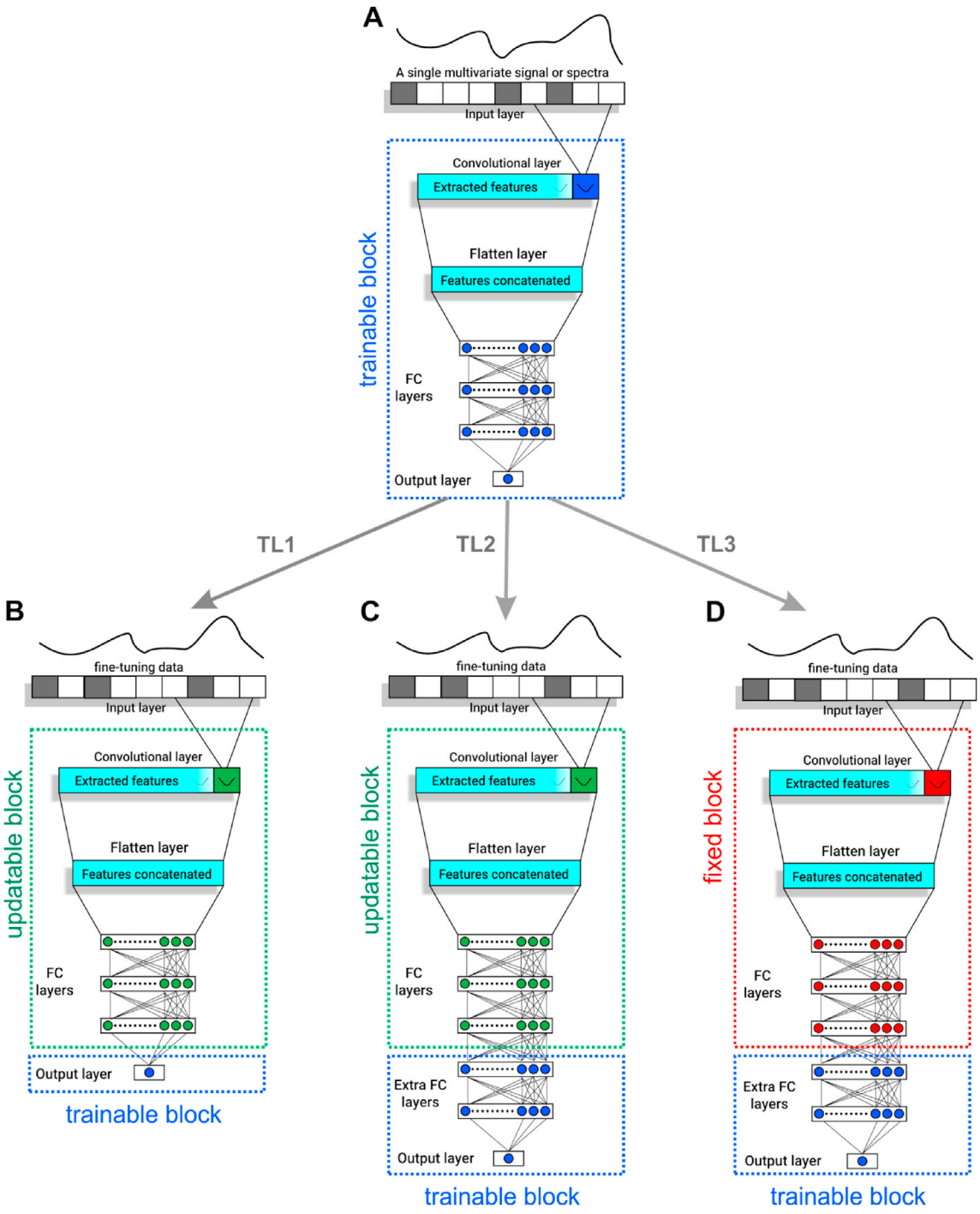

Fig. 1. A summary of deep learning (DL) and transfer learning (TL) architecture for model training and transfer. (A) The DL architecture for primary model training, (B) the model used in TL1, (C) the model used in TL2 and (D) the model used in TL3. Blue represents units that are initialized using 'He_normal' distribution and allowed normally training; green represents units initialized with parameters from the pre-computed base model and are updated using fine-tuning data; red represent units with fixed ("untrainable") parameters loaded from the pre-computed base model. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

adapted along with the optimization with an iterative algorithm (using tf.keras.ReduceLROnPlateau()function) that dynamically decreased the LR to ensure that the best convergence behaviour is achieved. The loss function used was the mean squared error (MSE) accompanied by an L2 penalty (β) on the model weights (layer regularization). L2 regularization in DL is particularly important to alleviate the overfitting that CNNs tend to exhibit. During the model optimization phase, the maximum number of training epochs was set to 700. To avoid overfitting, an 'Early Stopping' approach was used with the tf.keras.EarlyStopping() implementation. The training stop criterion was the improvement of at least 10^{-4} in the validation/tuning loss after 100 consecutive epochs.

The model was optimized using a grid search approach over 3 of the hyper-parameters of the CNN: the "kernel width" on the convolutional layer ∈ [, 10, 15, 20, 25, 30], the number of samples for each training

batch, "batch size" ∈ [32, 64, 128, 192, 256, 384, 512] and the strength of the L2 regularization, β ∈ [0.001, 0.003, 0.008, 0.01, 0.015, 0.02, 0.03, 0.07, 0.1, 0.3]. The optimization process was done separately for the mango and melamine data. To define the best model, the hyper-parameters choice was done based on the trade-off between lowering the overfitting on the calibration set and keeping the RMSE on the tuning set as low as possible. The last step before using the data sets was to perform a column-wise (features) standardization to rescale the data into a range more suitable to the CNN.

2.3. Transfer learning

Once the optimal initial CNNs for mango and melamine cases were obtained, the next step was to update them to adapt for the new unseen variability. The new variability in the case of mango was the new seasonal variability in mango fruit, while in the case of melamine, the variability was related to the change in the recipe of the raw material. To update the model, transfer learning (TL) was tentatively implemented in the three separate ways described below:

1. In the first approach, the CNN layer's original architecture (Fig. 1A) does not suffer any modification**.** In the convolutional and FC layers, the model's parameters were initialized with pre-trained weights and biases values and they could retrain/update further based on the fine-tuning data (Fig. 1B). The parameters on the output layer were reinitialized and trained from scratch. In the following part of manuscript, this approach will be called **TL1.**
2. In the second approach, the CNN architecture was augmented with two extra FC layers (36 neurons followed by an 'elu' activation each) just before the output layer, (Fig. 1C). The parameters in the original convolutional and FC layers were initialized with pre-trained values and further trained (updated). The weights in the output and in the two newly added FC layers were initialized using 'He_normal' and trained on the fine-tuning data. In the following part of manuscript, this approach will be called **TL2.**
3. In the third approach (Fig. 1D), the CNN architecture and the model weights were kept fixed with their pre-trained values (not updated) and, to learn the new variability in the fine-tune data, only the two extra FC layers and output layer could be trained. This approach will be called **TL3.**

Once the models were updated through TL, their performance was tested on the independent test set. TL was implemented separately for mango and melamine case.

2.4. Baseline comparison

To have a comparison of the proposed TL approaches, as a baseline, three different existing model updating methods were used. The first method was a PLS model [2,3] recalibrated by combining the training and fine-tuning data (a subset of new scenario) [31]. The second was the recently proposed semi-supervised method to PFCE. The semi-supervised PFCE framework uses the fine-tuning data (a subset of new scenario) and imposes correlation constraints to update the existing PLS model [32]. The third was a recently proposed DL method that enables sharing weights between 1-D convolutional neural networks [39]. The weight sharing network was implemented with the same settings as designed in Ref. [39]. The weight sharing framework allows training a big and a small data set jointly with two CNN models such that the model trained on small data can learn from the patterns extracted by the model being trained on big data. In this case, the training set was considered as the big data and the fine-tuning set was considered the small data. In all the cases, the models were tested in the same test data as used for TL testing. The performance of all models tested on the final test set was assessed by the root mean squared error of prediction (RMSEP).

3. Results and discussion

3.1. Primary CNN for mango and melamine cases

To reach the best primary models for CNN, three main hyper-parameters i.e., L2 regularization (β), batch size and kernel width were optimized. From the primary exploration of the CNN architecture with the mango and melamine data sets, β was found to have the most impact on the overfitting of the models. For example, in Fig. 2, models corresponding to β = 0.001 (solid lines) and β = 0.1 (dashed lines) for melamine data set, with varying batch sizes and kernel widths are presented. It can be noted that the higher value of β = 0.1 decreased the difference between the calibration and tuning RMSE sets, i.e., decreased the overfitting on the calibration set. On the other hand, the varying batch sizes and kernel widths showed stable RMSE for both β = 0.001 (solid lines) and β = 0.1 (dashed lines).

To choose the best L2 regularization β hyper-parameter, for each value, the mean of the difference between tuning and calibration RMSE (a measure of overfitting), and the mean of the tuning RMSE was used. The averaging was done over the models that were computed in the grid-search using the remaining combinations of "kernel widths" and "batch sizes". After that, based on their profiles (Figs. 3, 4), the lowest interception points were chosen corresponding to the optimal models. For the melamine data (Fig. 3), $\beta = 0.07$ provides the best compromise between low overfitting and low tuning RMSE. For the mango data (Fig. 4), $\beta = 0.008$ was chosen. The same rationale is applied to the other two hyper-parameters although their influence in the RMSE is smaller. Both data sets behaved differently in hyper-parameter space (Figs. 3 and 4), and because of that, setting up a general method for choosing the best model was difficult. For both data sets, the used criteria chose points that were not coincident with the minima of the tuning RMSE (Figs. 3A and 4A). This is also a frequent problem in standard chemometrics where the most robust models are not always the ones that produce the lowest metrics on the tuning set [40].

The best models based on the parameters found in Figs. 3 and 4 reached the RMSE of 1.24 °C for turbidity point prediction during melamine production, and 0.75% for predicting DM in mango fruit, on the tuning sets. However, the RMSEP increased when the models were tested in the new scenario, for both mango and melamine cases. A summary of the CNN models made on the primary batch (train data) and tested on the data from the new scenarios are shown in Fig. 5. It can be noted that for the DM prediction in mango, the RMSEP was increased from 0.75 to 0.89%, and for the melamine case, the RMSEP for turbidity point prediction increased from 1.24 °C to 1.92 °C. Such an increase in RMSEPs with the use of primary CNN models on the new scenario data, confirms the need for the model update.

3.2. Transfer learning and models re-calibration

3.2.1. Mango data

To this end, the optimized CNN models for DM and turbidity point prediction in mango and melamine process were chosen. Furthermore, it was shown that the prediction performance of these primary models degraded as they were applied to data of a new scenario (Fig. 5), as the RMSE was higher compared to the tuning set. Hence, the models were updated by different TL approaches to account for the new variability. A summary of the results for mango and melamine cases is shown in Figs. 6 and 7, respectively. In the case of mango, the lowest RMSEP (0.58%) was reached with the TL2 approach (Fig. 6F) proposed in this study, i.e., retrain the weights of the initial model (from pre-trained values) and adding extra FC layers to the original architecture. The PLS model recalibrated by adding new season data performed the poorest with RMSEP of 0.95% (Fig. 6A). This is indicative that the simple recalibration of a PLS model by incorporating new data is not an ideal solution. Even a new PLS model (Fig. 6B) fitted to a small set of data from a new season performed better than the recalibrated PLS model (Fig. 6A). The semi-

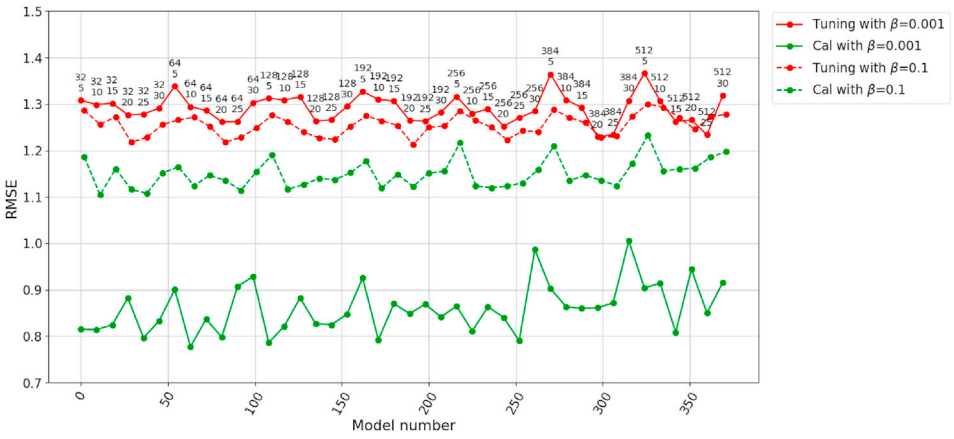

Fig. 2. The effect of β, batch size and kernel width on the performance of DL models for melamine data set. Dashed lines stand for the computed RMSE of the calibration (green) and tuning (red) set for models with β = 0.1 across multiple batches and kernels widths (top and bottom label numbers). Solid lines stand for the computed RMSE of the calibration (green) and tuning (red) set for models with β = 0.001 across multiple batches and kernels widths (top and bottom label numbers). . (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

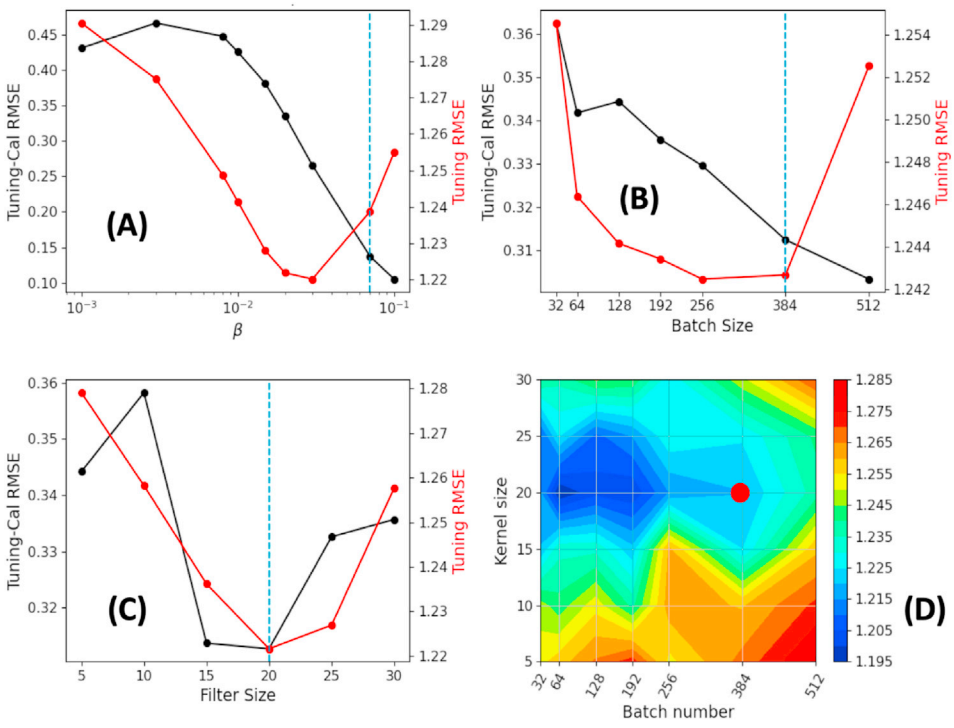

Fig. 3. Hyper-parameters choice for the melamine data. (A) tuning minus calibration RMSE (black) and tuning RMSE (red) vs *β, (B) tuning minus calibration RMSE and tuning RMSE vs batch size, (C) tuning minus calibration RMSE and tuning RMSE vs filter size. Vertical dashed lines stand for the chosen hyper-parameter. (D) shows RMSE for the tuning set for models with β = 0.07 as a function of batch number and kernel size. The red dot shows the chosen hyper-parameters (batch=384, kernel size=20).* . (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

supervised PFCE approach was able to update the PLS model made on primary data with the small set of new season data but obtained a RMSEP of 0. 77%. That was like the RMSEP obtained with new PLS performed on the new season data alone. Hence, the PFCE did not prove that it gained anything from the variability present in the old model when updated with the new season data. The weight share DL (Fig. 6D), and TL1 approach proposed in this study (Fig. 6E), performed similar and attained RMSEPs of 0.60% and 0.61%, respectively. The TL3 approach (Fig. 6G) proposed in this study performed the poorest compared to the TL1 (Fig. 6E) and TL2 (Fig. 6F) approaches.

3.2.2. *Melamine data*

In the case of the melamine production process, the lowest RMSEP (1.47°C) was attained with the TL1 and TL2 approaches (Fig. 7E and F) proposed in this study. Like the mango case, the PLS model recalibrated by adding new recipe data performed the poorest with RMSEP of 2.46°C (Fig. 7A). A new PLS model (Fig. 7B) calibrated with a small set of data from the new recipe performed better than the recalibrated PLS model (Fig. 7A). The semi-supervised PFCE approach was able to update the PLS model made on primary data and attained RMSEP of 1.72°C which was lower than both the recalibrated PLS, and the new PLS performed on the new recipe data alone. Hence, the PFCE gained from the variability present in the old model as well the new variability in the data from the new recipe. The weight share DL (Fig. 7D) performed poorer than the PFCE approach and reached a RMSEP of 1.96°C. For this data set, the TL3 approach (Fig. 7G) proposed in this study also performed worse when compared to the TL1 (Fig. 7E) and TL2 (Fig. 7F) approaches. However, even this least efficient approach (TL3) outperformed the PFCE and weight share DL approach.

In both the mango and melamine cases, the PLS recalibration approach performed the poorest of all the presented approaches. Nonetheless, in the spectral data analysis domain, PLS recalibration is a widespread practice where recalibration of the PLS model is performed by incorporating some new data [31]. A probable reason for this is that it is a simple approach which can be implemented in all software resources available for chemometric analysis. However, in terms of classical chemometrics frameworks, it is advised that future users search for better alternatives and explore recently developed dedicated model updating

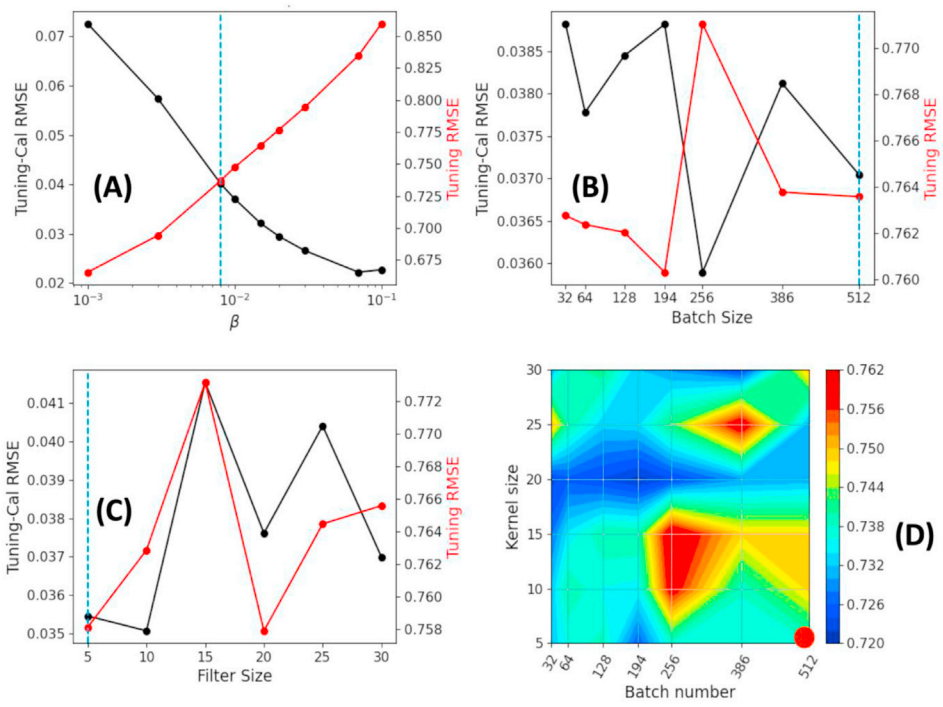

Fig. 4. Hyper-parameter choice for the mango data. (A) tuning minus calibration RMSE (black) and tuning RMSE (red) vs β, *(B) tuning minus calibration RMSE and tuning RMSE vs batch size, (C) tuning minus calibration RMSE and tuning RMSE vs filter size, and (D) RMSE for the tuning set for models with* $\beta = 0.008$ *as a function of batch number and kernel size. The red dot shows the chosen hyper-parameters (batch=512, kernel size=5).* . (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

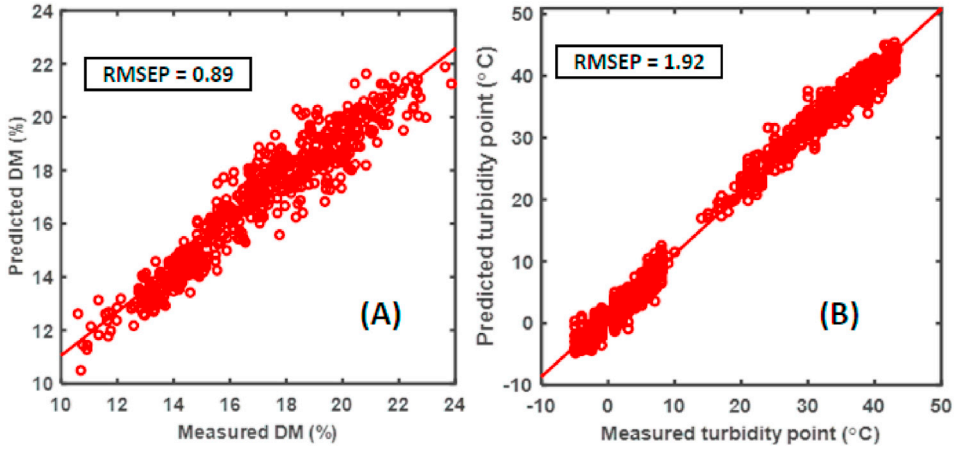

Fig. 5. A summary of convolutional neural networks (CNNs) models tested on a new scenario for mango and melamine case. (A) Primary CNN model for DM prediction in mango tested on a new season data, and (B) Primary CNN model for turbidity point measurement on melamine production process tested on a new recipe of melamine production.

approaches such as semi-supervised PFCE. In this study, the PFCE showed its superiority to PLS recalibration by reaching lower RMSEP from the same data set. Further, the PFCE has the main benefit that it is a parameter free framework, and just requires the regression vector of the existing PLS model to be updated. Although PFCE was better than the PLS recalibration, its performance was unmatchable to the TL approaches based on DL models proposed in this study. The main reason was that the foundation of the TL was the non-linear properties of the CNN which already outperformed the PLS analysis, the foundation of the PFCE update procedure. In this study, three different methods of implementing transfer learning were proposed (TL1, TL2 and TL3). TL1 and Tl2 shared a similar foundation as they both allowed the primary model parameters (weights and biases) to be updated based on the data from the new scenario, while the TL3 only allowed the new FC layers to learn the latest information. For both cases presented in this study i.e. mango and melamine, a better performance of the TL1 and TL2 approaches compared to the TL3 was noted. Hence, based the results from this study we suggest that future users should keep the initial CNN layer architecture, initialized with the pre-trained weights, trainable, and explore the model update with (TL2) or without extra layers (TL1).

4. Conclusions

This study, for the first time, showed three approaches to update the CNN models to make them suitable to be used in a new scenario where spectral data re-calibration was needed. Out of the three approaches, two approaches allowed the old model parameters to be retrained, while one approach kept the parameters of the old model intact and only learned with the newly added layers. The results showed that letting the old model parameters to readjust led to better predictive performance for both the cases presented i.e., mango fruit quality and melamine production process. Further, having extra FC layers over the trainable base model improved the predictive performance (compared to no added layers) for DM prediction in mango fruit. The recalibration of PLS models

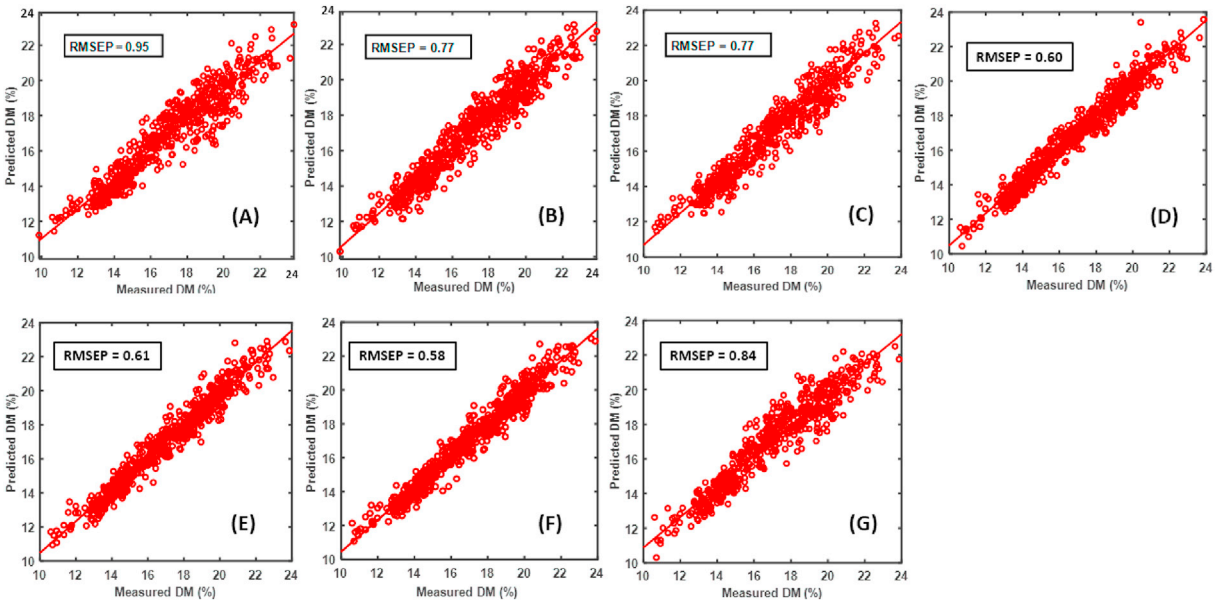

Fig. 6. A summary of performance of different modelling approaches on the data from new season of mango fruit. (A) Recalibrated PLS model by combining some data from new season, (B) PLS model made on data from new season, (C) semi-supervised parameter free calibration enhancement, (D) DL with weight sharing framework, (E) DL model update with TL1 approach, (F) DL model update with TL2 approach, and (G) DL model update with TL3 approach.

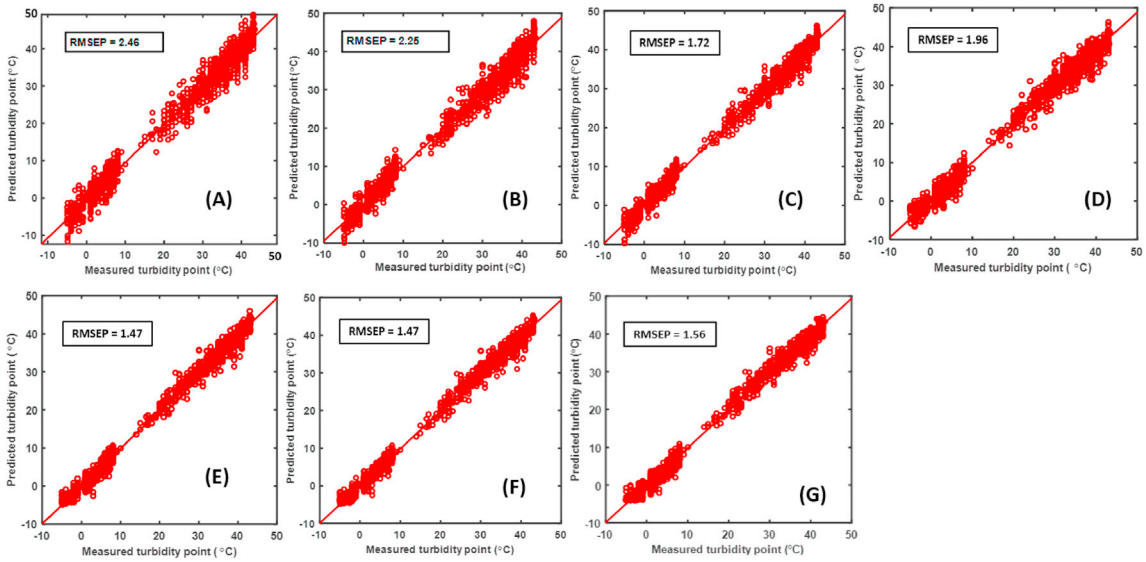

Fig. 7. A summary of performance of different modelling approaches on the data from new recipe for melamine production. (A) Recalibrated PLS model by combining some data from new season, (B) PLS model made on data from new season, (C) semi-supervised parameter free calibration enhancement, (D) DL with weight sharing framework, (E) DL model update with TL1 approach, (F) DL model update with TL2 approach, and (G) DL model update with TL3 approach.

was found to be the worst approach for updating calibration models to be used in a new scenario. Out of all the model updating approaches presented, the TL approach was the best in terms of reaching the lowest RMSEP for both the mango and the melamine production process data set. In summary, TL supplies a useful and successful method of updating spectral models surpassing even some of the most recently established methodologies in classical chemometrics. However, some hard challenges still need to be overcome before DL models receive widespread acceptance in the chemometric community. One is related to the need for more research about the optimal NN architectures to use with spectral data. This requires a cross interbreed between experts in the field of DL and the field of chemometrics. Another challenge is the widespread use of DL models implementations in the Python programming language, as most of the chemometric community is currently based on either MATLAB or R programming language. This is bound to change since both the later languages have already started to incorporate DL packages in their distributions.

Author statement

Puneet Mishra: Conceptualization; Methodology; Software; Writing - Original Draft; Data Curation.

Dário Passos: Software; Formal analysis; Writing - Review & Editing.

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] S. Wold, C. Albano, W.J. Dunn, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg, M. Sjöström, Multivariate data analysis in chemistry, in: B.R. Kowalski (Ed.), Chemometrics: Mathematics and Statistics in Chemistry, Springer Netherlands, Dordrecht, 1984, pp. 17–95.

[2] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometr. Intell. Lab. Syst. 58 (2001) 109–130.

[3] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, Anal. Chim. Acta 185 (1986) 1–17.

[4] M. Daszykowski, S. Serneels, K. Kaczmarek, P. Van Espen, C. Croux, B. Walczak, Tomcat, A MATLAB toolbox for multivariate calibration techniques, Chemometr. Intell. Lab. Syst. 85 (2007) 269–277.

[5] S. Wold, PLS Modeling with Latent Variables in Two or More Dimensions, 1987.

[6] C. Cui, T. Fearn, Modern practical convolutional neural networks for multivariate regression: applications to NIR calibration, Chemometr. Intell. Lab. Syst. 182 (2018) 9–20.

[7] 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, Chemometr. Intell. Lab. Syst. 200 (2020) 103996.

[8] 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.

[9] 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, Chemometr. Intell. Lab. Syst. 172 (2018) 188–193.

[10] S. Malek, F. Melgani, Y. Bazi, One-dimensional convolutional neural networks for spectroscopic signal regression, J. Chemometr. 32 (2018), e2977.

[11] L.M. Yuan, F. Mao, X.J. Chen, L.M. Li, G.Z. Huang, Non-invasive measurements of 'Yunhe' pears by vis-NIRS technology coupled with deviation fusion modeling approach, Postharvest Biol. Technol. 160 (2020).

[12] K.B. Walsh, V.A. McGlone, D.H. Han, The uses of near infra-red spectroscopy in postharvest decision support: a review, Postharvest Biol. Technol. 163 (2020) 111139.

[13] K.B. Walsh, J. Blasco, M. Zude-Sasse, X. Sun, Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: the science behind three decades of commercial use, Postharvest Biol. Technol. 168 (2020) 111246.

[14] N.T. Anderson, K.B. Walsh, J.R. Flynn, J.P. Walsh, Achieving robustness across season, location and cultivar for a NIRS model for intact mango fruit dry matter content. II. Local PLS and nonlinear models, Postharvest Biol. Technol. 171 (2021) 111358.

[15] N.T. Anderson, K.B. Walsh, P.P. Subedi, C.H. Hayes, Achieving robustness across season, location and cultivar for a NIRS model for intact mango fruit dry matter content, Postharvest Biol. Technol. 168 (2020) 111202.

[16] P. Mishra, R. 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. (2020) 103547.

[17] P. Mishra, J.M. Roger, F. Marini, A. Biancolillo, D.N. Rutledge, Fruitnir-Gui, A graphical user interface for correcting external influences in multi-batch near infrared experiments related to fruit quality prediction, Postharvest Biol. Technol. (2020) 111414.

[18] 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.

[19] M. Zeaiter, J.M. Roger, V. Bellon-Maurel, Dynamic orthogonal projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR-based monitoring of wine fermentations, Chemometr. Intell. Lab. Syst. 80 (2006) 227–235.

[20] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, C. Liu, A Survey on Deep Transfer Learning, Springer, pp. 270-279.

[21] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (2017) 84–90.

[22] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 arXiv preprint arXiv:1409.1556.

[23] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 770–778.

[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going Deeper with Convolutions, pp. 1-9.

[25] B. Espejo-Garcia, N. Mylonas, L. Athanasakos, S. Fountas, I. Vasilakoglou, Towards weeds identification assistance through transfer learning, Comput. Electron. Agric. 171 (2020) 105306.

[26] A. Abdalla, H. Cen, L. Wan, R. Rashid, H. Weng, W. Zhou, Y. He, Fine-tuning convolutional neural network with transfer learning for semantic segmentation of ground-level oilseed rape images in a field with high weed pressure, Comput. Electron. Agric. 167 (2019) 105091.

[27] 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.

[28] 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.

[29] D. Zhang, Y. Ding, P. Chen, X. Zhang, Z. Pan, D. Liang, Automatic extraction of wheat lodging area based on transfer learning method and deeplabv3+ network, Comput. Electron. Agric. 179 (2020) 105845.

[30] 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.

[31] 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.

[32] J. Zhang, B. Li, Y. Hu, L. Zhou, G. Wang, G. Guo, Q. Zhang, S. Lei, A. Zhang, A parameter-free framework for calibration enhancement of near-infrared spectroscopy based on correlation constraint, Anal. Chim. Acta 1142 (2021) 169–178.

[33] J.S. Larsen, L. Clemmensen, Deep Learning for Chemometric and Non-translational Data, 2019 arXiv preprint arXiv:1910.00391.

[34] R.W. Kennard, L.A. Stone, Computer aided design of experiments, Technometrics 11 (1969) 137–148.

[35] R. Nikzad-Langerodi, F. Sobieczky, Graph-based Calibration Transfer, 2020 arXiv preprint arXiv:2006.00089.

[36] R. Nikzad-Langerodi, W. Zellinger, S. Saminger-Platz, B.A. Moser, Domain adaptation for regression under Beer–Lambert's law, Knowl. Base Syst. 210 (2020) 106447.

[37] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, pp. 770-778.

[38] D.P. Kingma, J. Ba, Adam, A Method for Stochastic Optimization, 2014 arXiv preprint arXiv:1412.6980.

[39] J.S. Larsen, L. Clemmensen, Weight Sharing and Deep Learning for Spectral Data, ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP), 2020, pp. 4227–4231.

[40] F. Westad, F. Marini, Validation of chemometric models – a tutorial, Anal. Chim. Acta 893 (2015) 14–24.

[41] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359.

[42] Karl Weiss, Taghi M. Khoshgoftaar, DingDing Wang, A survey of transfer learning, J Big Data 3 (9) (2016) 1–40.

[43] F. Zhuang, et al., A comprehensive survey on transfer learning, Proc. IEEE 109 (1) (Jan. 2021) 43–76, https://doi.org/10.1109/JPROC.2020.3004555.