Received: 11 April 2021 | Revised: 28 June 2021 | Accepted: 6 July 2021

DOI: 10.1002/cem.3367

RESEARCH ARTICLE

Journal of CHEMOMETRICS WILEY

Deep chemometrics: Validation and transfer of a global deep near-infrared fruit model to use it on a new portable instrument

Puneet Mishra[1] | **Dário Passos**[2]

[1]Wageningen Food and Biobased Research, Wageningen University and Research, Wageningen, The Netherlands

[2]CEOT | Physics Department, Universidade do Algarve, Faro, Portugal

Correspondence
Puneet Mishra, Wageningen Food and Biobased Research, Wageningen University and Research, Bornse Weilanden 9, P.O. Box 17, 6700AA, Wageningen, The Netherlands.
Email: puneet.mishra@wur.nl

Abstract

Recently, a large near-infrared spectroscopy data set for mango fruit quality assessment was made available online. Based on that data, a deep learning (DL) model outperformed all major chemometrics and machine learning approaches. However, in earlier studies, the model validation was limited to the test set from the same data set which was measured with the same instrument on samples from a similar origin. From a DL perspective, once a model is trained it is expected to generalise well when applied to a new batch of data. Hence, this study aims to validate the generalisability performance of the earlier developed DL model related to DM prediction in mango on a different test set measured in a local laboratory setting, with a different instrument. At first, the performance of the old DL model was presented. Later, a new DL model was crafted to cover the seasonal variability related to fruit harvest season. Finally, a DL model transfer method was performed to use the model on a new instrument. The direct application of the old DL model led to a higher error compared to the PLS model. However, the performance of the DL model was improved drastically when it was tuned to cover the seasonal variability. The updated DL model performed the best compared to the implementation of a new PLS model or updating the existing PLS model. A final root-mean-square error prediction (RMSEP) of 0.518% was reached. This result supports that, in the availability of large data sets, DL modelling can outperform chemometrics approaches.

KEYWORDS

artificial intelligence, calibration transfer, deep learning, fruit-chemistry, spectroscopy

1 | INTRODUCTION

Rapid prediction of fruit quality is of wide interest to access the ripeness and 'readiness to eat' stage of fresh fruit.[1–4] Such a rapid prediction of fruit quality relies on decision making at several stages of the fresh fruit supply chain, such as the best date for harvest and monitoring of fruit during controlled ripening or storage.[5,6] Furthermore,

non-destructive prediction of fruit quality reduces the experimental time and costs as well as it minimises the food losses along the supply chain.[5,6] A popular non-destructive technique for rapid analysis of fresh fruit traits is Vis–NIR portable spectroscopy.[6–8] Vis-NIR spectroscopy explores the interaction of Vis-NIR electromagnetic radiation and their correlation with several chemical and physical properties of fruit.[7,8] Particularly, the visible (Vis) part of the spectrum explains mainly the fruit skin pigments which in some cases are correlated with of fruit ripeness stage, while the near-infrared (NIR) part captures the signal related to the chemical components[9,10] on the pulp such as moisture and sugars which are a good proxy for fruit maturity.[6,11,12] In some works, Vis-NIR spectroscopy has also been used to explain physical properties such as firmness which is a good indicator of ripeness levels of fruit such as avocados.[13]

Motivated by the reliable performance of portable Vis-NIR spectroscopy to estimate several physicochemical traits in fruit, several user-friendly low-cost commercial portable spectrometers have appeared in the market.[5,6] The portability of spectrometers ranges from mobile phone connected pocket spectrometers[14] to handheld spectrometers with embedded computing systems.[5,11,12] Apart from the market level readiness of the hardware, the main challenge related to the distributed calibration of the spectrometer persists.[15,16] This is because Vis-NIR spectrometers are highly dependent on the pre-calibration of the sensors before their deployment in practice.[7] At present, in most of the practical cases, the users must carry out their calibrations based on the application of interest by measuring a few sets of new samples.[5–7] These calibrations also suffer failure when met with samples having unmodeled variability.[17–19] For fresh fruit, the model failure is commonly due to a high biological variability.[20,21] Biological variability can be related to different cultivars, the season of harvest, storage conditions and ripening stages of fruit.[5,6,20,21] Hence, a natural solution to deal with this calibration failure problem is to measure a wide range of samples from different cultivars, seasons of harvest, storage conditions and ripening stages to calibrate global models.[5,6] These global models are expected to work well in the face of the unseen variability that is expected to be met when predicting a new sample.[6] Measurement of samples covering a wide variability is the first step to achieve global models; however, the second step is to properly model the data to develop robust models that not only learn the relationship between spectra and the property of interest but also learn the complex pattern such as the causes for seasonal, cultivar and ripeness related variabilities.[22,23] Several authors have widely tried linear, non-linear, global and local modelling approaches to achieve robust models and have achieved satisfactory performance.[6,22,23]

A key feature of portable spectroscopy is its wide-scale usage compared to the scientific laboratory-based spectrometers.[14,15,24] Such a wide-scale usage has an advantage as it can generate enormous information-rich data sets. Such large data sets covering a wide variety of samples can allow development of robust models, otherwise limited to the data generated by a single user. In fruit spectroscopy, such data sets are gaining popularity. For example, recently a mango data set having 11,691 spectra and reference dry matter (DM) measurements was made publicly available.[25] The data set was explored in several studies with traditional chemometric and machine learning approaches and the best root-mean-square error prediction (RMSEP) was noted as 0.84%.[22,23] Taking advantage of the size of such a large data set, recently a deep learning model was proposed.[26] That DL model outperformed all the earlier linear, non-linear, local, and global modelling approaches and reached a RMSEP of 0.79% on the same test set reported in earlier studies. The performance of the DL model was improved by removing remaining outliers and a RMSEP of 0.76% was reached.

A DL model trained on a large data set as well with the best predictive performance is expected to be used in practice. For example, users who bought a new portable spectrometer and interested in predicting DM in mango will be highly interested to use the existing DL model covering wide variability of samples, instead of recalibrating the sensors from scratch. In earlier works related to DL modelling of the mango data set,[26] the main aim was to optimise and reach a robust DL model which improves on the RMSEP previously reported on the similar data set. In that regard, to have a fair comparison, the DL model in the earlier study[26] was tested on the predefined test set available in the mango data set, that is, data corresponding to the harvest season from 2018.[22,23] However, in a practical scenario, a new user will be interested in using the DL model in a different instrument compared to the instrument used for measurement of the mango data set. Hence, currently, it is unclear if the DL model developed in Mishra and Passos[26] can be directly used on a new-instrument or if it requires some form of model adaptation or recalibration before it can be used with the new instrument. Furthermore, since the DL model was trained on the samples from the harvest of the year 2015, 2016 and 2017, it is important to explore if the model performs well when applied to samples from a new harvest season measured on a different spectrometer. To the best of our literature search and experience with the development of the primary DL model,[26] this work is the first to validate the mango DL model[26] for its use on a new instrument and on mango samples of harvest from a new season, that is, the year 2020.

The study has three main aims:

1. To provide independent validation of earlier DL model[26] related to mango DM prediction on a new test set measured in a local laboratory setting with a new spectral instrument.
2. To retrain/optimise the model concerning the season variability to make it robust to seasonal differences.
3. Since the measurements in this study were performed with a different spectrometer, the DL model was updated with transfer learning (TL)[27] to compensate for the instrument differences. The TL approach can be assumed as the calibration transfer approach for DL models just as calibration transfer between spectrometer is required for standard chemometric calibrations.[28]

2 | MATERIALS AND METHODS

2.1 | Data sets

Two data sets were used for modelling and independent testing. The modelling data set was the freely available mango data set, and the independent testing data set was generated experimentally for this study. More details on data sets are as follow.

2.1.1 | Open-access mango data set

The first data set has spectral and reference DM measurements on mango from 4 harvest seasons 2015, 2016, 2017 and 2018.[25] Data were acquired with a Portable F750 instrument (Felix Instruments, Camas, USA). Drying for DM measurements was performed with an oven dryer. Detailed information on data set can be assessed at previous works,[22,23] and the data can be assessed at Anderson et al.[25] At the online data source,[25] there are total 11,691 spectra, which are pre-partitioned as calibration (10,243 from year 2015, 2016 and 2017) and independent test set (1448 from year 2018). The data spectral range was reduced to 684–990 nm (103 variables) to keep the analysis comparable to earlier work.[26] The original data has several outliers in the calibration set, hence, in this study, a more stringent outlier removed version of data was used as suggested in Mishra and Passos.[26] Furthermore, based on the suggestion of this recent study, prior to any analysis the data were augmented with spectral pre-processing.[26] The final data set has 9914 samples for model training and validation and 1448 samples for an independent test of the model. However, after the data augmentation suggested in Mishra and Passos,[26] the total number of variables were 618 by the stacking of five pre-processed data blocks to the original absorbance data block. The five different pre-processing's were 1st derivative and 2nd derivative with Savitzky–Golay[29] filter with window size 13 and polynomial order of 2, standard normal variate (SNV),[30] SNV + 1st derivative and SNV + 2nd derivative.

2.1.2 | New mango data set

To perform an independent validation of the DL model proposed in Mishra and Passos,[26] and to adapt it for future use on a new instrument, new mango samples were measured with a new portable spectrometer. The mango samples were from a new harvest, that is, the year 2020, and have the origin in Brazil. Furthermore, the mangoes were of two cultivars, that is, 'Keitt' and 'Kent'. A total of 540 (270 'Kiett' and 270 'Kent') mangoes were bought from a local vendor in The Netherlands. The spectrometer was a F750 spectrometer (Felix Instruments, Camas, USA). The spectrometer used was a similar but not the same instrument as used for the generation of open mango data set. Since most fruit were hard green at the time of purchase, to create a wide variation in the ripeness level, the mangoes were stored in different temperatures, that is, 12°C, 17°C and 23°C at 85% RH (Relative Humidity). The low-temperature storage hinders the fruit ripening while hot temperature storage accelerates the ripening. 90 'Kent' and 90 'Keitt' mangoes were stored at each temperature level. Mangoes were stored for 1 week in the fruit ripening facilities and before the experiment day was stored for 1 day in the same temperature condition (23°C at 85% RH) to reduce any effect of temperature on the spectral measurements. On the day, the experiment took place, the spectral measurements were performed first and after that, a part of mango fruit was collected from the spot of spectral measurement. The peel

of the part was removed and later the fresh weight of the fruit's flesh was measured using an electronic balance and dried in a hot-air oven. After drying, the dried fruit was measured, and the DM was estimated and expressed in %. In total, 540 spectral and DM reference measurements were generated. The spectra were at first trimmed to match the spectral range (684–990 nm total 103 variables) as used for the development of DL model. Later, outlier removal was performed using the T^2 and Q statistics plots obtained with the PLS decomposition of spectra with the DM measurements. The outlier's removal was performed interactively by plotting T^2 and Q statistics plots and later dropping any observed outlying samples. Later, the spectral data were augmented as suggested in Mishra and Passos,[26] that is, by pre-processing the same data with five different pre-processing approaches. After outlier removal and data augmentation, the total remaining samples were 510, each having 618 variables. All data analysis was performed with the augmented data as needed for the application of the old DL model[26] as well due to a improved performance of the model on the augmented data compared to raw data.[26]

2.2 | Data modelling

In this study, there were three main directions for data modelling. The first was the direct application of the DL and PLS[26] models developed on the open-access mango data set[25] to the new mango data set measured in this study. The first part aimed to study if the previously developed DL and PLS models[26] generalise well to data from a new harvest measured with a different instrument. The second was the development of new DL and PLS models by combining all the data of the open-access mango data and testing it on the new mango data set. A key point to note is that in the earlier study, the DL model was based on data from seasons 2015, 2016 and 2017 and the data from season 2018 was used as independent test set. In this study, the second part aimed to exploit the full potential of the open-access mango data set, hence, modelling was performed by combining data from all four seasons, that is, 2015, 2016, 2017 and 2018. The third part was related to the adaption and update of the DL model to compensate for the differences in the instrument. The third part was necessary as the primary DL model was made on the open-access mango data where the spectral measurements were performed with a similar but not the same spectrometer. Hence, it was a classic case of calibration transfer but without standards as the two spectrometers belongs to different scientific research groups based on different continents. All models were implemented using the Python (3.6) language and the Tensorflow 2.1 framework.

2.2.1 | Direct application of old models on a new independent test set

In Mishra and Passos,[26] DL and PLS models based on open-access mango data set were proposed. The models were based on the calibration set of the open-access mango data set. Previously, the models were tested on the test set of the open-access mango data only which corresponds to the mango of the year 2018 harvest. However, in this study to study their generalisability, the DL and PLS models were tested on the new independent test set measured. The DL model was applied by defining the model architecture (1D CNN) and loading the pre-saved weights corresponding to the optimal DL model presented in Mishra and Passos.[26] The DL model was applied using the "model.predict()" function from TensorFlow/Keras. The application of PLS included the multiplication of the regression coefficients with the spectra from the new independent test set. The PLS was applied using the 'pls.predict()' function from the Scikit Learn library (https://scikit-learn.org/stable/). Prior to model application, data from the independent test set was scaled with respect to the calibration set of the open-access mango data set. The scaling was performed by estimating the SNV[30] for each variable of the independent test set with respect to the calibration set used for DL model training. In SNV,[30] at first, the means response is subtracted and later divided by the standard deviation. To scale the test data with respect to the calibration data set, the mean and standard deviation of calibration set were used.

2.2.2 | New deep learning model development

In the DL model presented in Mishra and Passos,[26] the modelling was limited to the calibration set as the data from the year 2018 harvest was used as the test set. However, since this study has a new independent test set, a new DL model can be made to learn variability related to seasonal differences. This can be achieved by using a calibration set of the open-access mango data for model training and the data from season 2018 as the validation/tuning set. Modelling in

such a way should improve the learning process associated to the variability of seasonal differences. Finally, the new DL model can be tested on the new independent test set measured in this study for a new season and with a new instrument.

To develop the new DL model, the same DL architecture as used in Mishra and Passos[26] and first proposed in Cui and Fearn,[31] was implemented. The architecture is composed of a five-layer convolutional neural network. The first layer of the network was a reshape layer which transform the spectra into a tensor, after that the tensor passes through a 1D-convolutional layer having a single filter, and later the convoluted signal is passed through three dense layers with 36, 18 and 12 neurons, respectively, and a final linear output layer with a neuron. Three key hyperparameters, that is, the width of the convolution filter (kernel width), the strength of the L2 regularisation (β) and the batch size, were optimised using a grid search approach. The grid search approach was used as presented in,[26] as it allows to sequentially optimise the hyperparameter and allows understanding the effect of different hyperparameters explored for model tuning. See Table 1.

Based on the possible combinations, a total 210 models were generated. To find the optimal model, an exploratory method used in the DL area was used. The hyperparameter that has a higher influence in the solution space is β (see supporting information for details). Therefore, a best β was chosen by finding the value that shows the smallest overfitting (the lower difference between validation and calibration RMSE) while keeping a low validation RMSE. Models prone to less overfitting tend to generalise better. Once β was found, the combinations of kernel and batch sizes were explored by analysing the calibration and validation RMSE maps. The strategy was to look for common minima (or common broader basins of attraction) in both maps to find which models belong in these intersection areas and discriminate between them by choosing the one(s) least affected by overfitting (the lowest difference between the RMSE of calibration and validation set). The final model was then tested on the new independent test set. As a baseline, a new PLS model was also developed by combining the calibration and validation sets of the open-access mango data. For PLS latent variables (LVs) optimisation a 10-fold cross-validation procedure was used The elbow point of the error plot was used for selecting the LVs.

2.2.3 | Standard free deep calibration transfer

The main difference between the open-access mango data set and the new test set measured in this study was the different instruments used for spectral measurements. Usually, to use a spectral model with data from a different instrument a calibration transfer step is needed before model application. In the case of DL modelling, calibration transfer can be performed by fine-tuning the weights of the old DL model using some new samples measured on the new instrument.[26] This recent article explains more broadly the TL approach as used in this study for calibration transfer.[27] The calibration transfer with TL is a standard free procedure to transfer DL models between instruments; hence, it does not require any measurements from the primary instrument and is well suited for this study as the primary instrument used to generate the open-access mango data was not available. To perform TL, the data generated in this study were divided into 60%/40% as fine-tune and external test set. The fine-tune set (60%) was used for TL, and later, the updated model was tested on the external test set (40%). As a baseline, a new PLS model was recalibrated combining the open-access mango data set and the fine-tune set. Furthermore, a PLS model solely based on the fine-tune data was also set up to see if updating the old model provided any benefit compared to developing a new PLS model with new data measured on the new instrument. All model performances were assessed using the RMSE. The RMSEP was used to access performance of all models.

3 | RESULTS AND DISCUSSION

A summary of DM distributions from different data sets is presented in Figure 1. The calibration set (Figure 1A) of the open-access mango data set used for the development of the primary DL model presented in Mishra and Passos.[26]

TABLE 1 Hyperparameters intervals used in the grid search optimisation

Hyperparameter	Interval
Batch size	32, 64, 128, 256, 512
Filter sizes	5, 10, 15, 20, 25, 30
L2 regularisation ($\boldsymbol{\beta}$)	0.001, 0.003, 0.008, 0.01, 0.015, 0.02, 0.03

Figure 1B shows the DM distribution of the test set of the open-access mango data set as used for model validation in Mishra and Passos.[26] Figure 1C shows the DM distribution of the independent test set measured in this study. The DM distribution of the independent test set has most of the samples with low DM (Figure 1C) compared to the DM range of the open-access mango data (Figure 1A,B). In earlier work, the DL model was tested on the test set of the open-access mango data set, which included mangoes with on average high DM (Figure 1B) compared to the DM distribution of calibration set (Figure 1A). Due to such a high DM, the DL model led to systematic higher RMSE on the test set of the open-access mango data set. In this study, the new measured samples have on average low DM (Figure 1C) compared to the DM distribution of calibration set (Figure 1A), how were in the similar DM range compared to the calibration set.

3.1 | Performance of old deep learning and PLS models

The performances of the direct application of the DL and PLS models presented in Mishra and Passos[26] are shown in Figure 2. At first, the performance of DL (Figure 2A) and PLS (Figure 2B) models is presented for the test set of the open-access mango data set. These results are the same as reported in Mishra and Passos,[26] and here presented to have a baseline comparison of the model with the testing performed on the new independent test set measured in this study. The DL model performance decreased on the independent test set compared to the earlier results reported for the test set of the open-access mango data set. The RMSE was increased from 0.787% to 1.053%. The PLS model, which previously performed worse, showed a lower RMSE of 0.754% on the independent test set. This showed that although the PLS model did not perform the best in the earlier study, it showed a better generalisation ability on a completely new test set measured on a completely different instrument. Furthermore, such an inferior performance of DL model shows that the DL model developed in Mishra and Passos[26] was sensitive to the instrument change compared to the PLS model presented in Mishra and Passos.[26] In this study, we considered two hypothesis for the inferior performance of the DL model on the new test set measured on a new instrument. The first hypothesis was that the new measurements were performed on different instrument compared to the measurements on which the primary DL model was developed. The instrument change plays a significant role in NIR spectroscopy and requires a pre-adjustment of the instrumental differences. The second hypothesis was that in earlier study,[26] a random data partition was used for DL model training. Random data partition can be inefficient in learning the seasonal variability and a better choice could be to retrain the model by using data from a complete year as the validation set. Both the limitations were dealt with in this study and the results were as follows.

3.2 | New deep learning model

Based on the open-access mango data set, a new DL model was trained with the calibration set and the test set for model tuning. Please note that the test set mentioned here was the data from the open-access mango data set and it was not the independent test set data measured in this study as used for the final evaluation of models. A key difference in the new model compared to the old model presented in Mishra and Passos,[26] was the use of the old test set for model

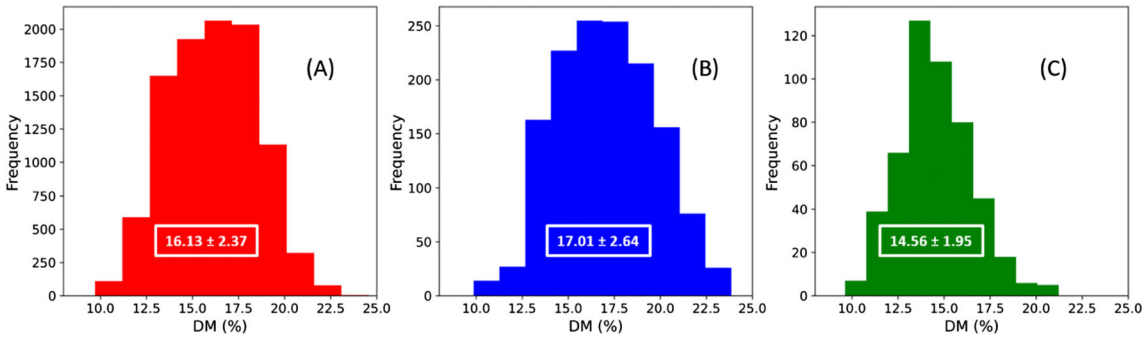

FIGURE 1 Distribution of dry matter in difference data sets. Open-access mango data set: (A) calibration set and (B) test set. (C) Mango data set measured in this study

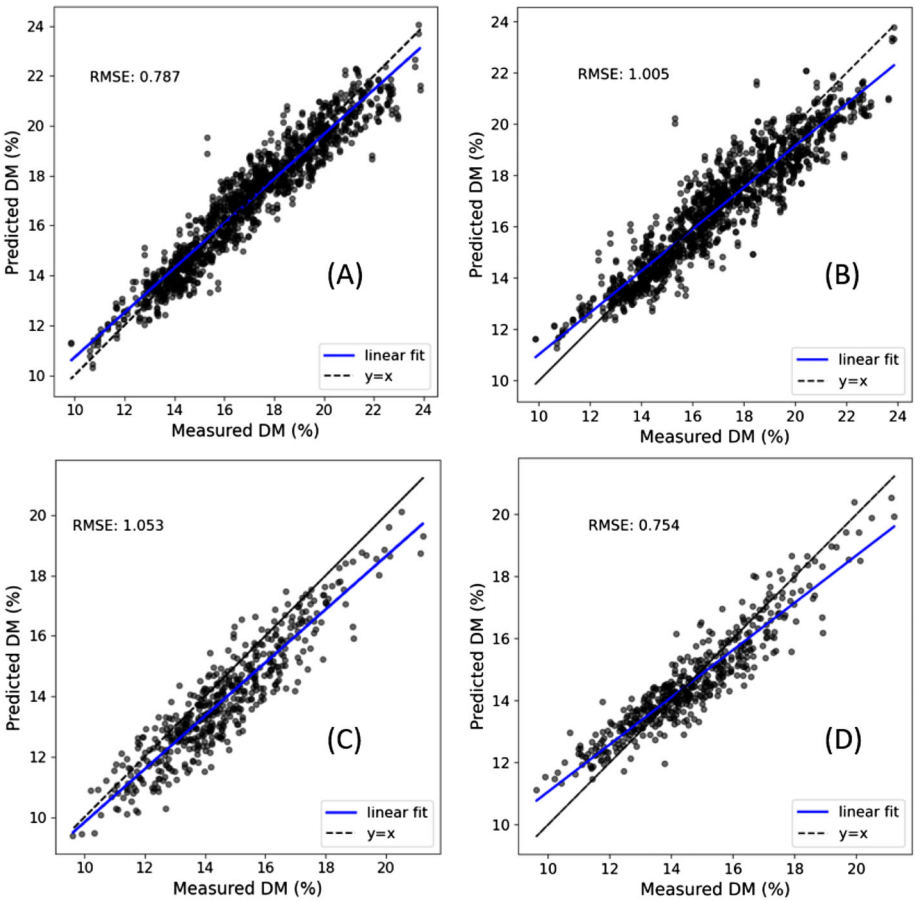

FIGURE 2 Model performances, (A) DL model developed in Mishra and Passos[26] and tested on the test of the open mango data set. (B) PLS model developed in Mishra and Passos[26] and tested on the test set of the open mango data set. (C) DL model developed for the open-access mango data[26] and tested on the new test set related to the mango harvest of year 2020. (D) PLS model developed for the open-access mango data[26] and tested on the new test set related to the mango harvest of year 2020

tuning. Since, the old test set of the open-access mango data set was from a new season of harvest, that is, the year 2018, using it for model tuning can allow the model to better capture the seasonal variabilities. It was hypothesised that the model tuned to learn seasonal variability could generalise well to a new season data, that is, the year 2020 harvest data measured in this study.

To reach the new DL model three different hyperparameters were optimised sequentially as explained in Section 2.2.2. Figure 3 shows the evolution of the mean RMSE of tuning and the difference between the means of RMSE (calibration and the tuning set) for different β. A $\beta = 0.02$ (dashed vertical line in Figure 3) was selected as optimal because it provides a good compromise between low overfitting (the difference between the mean RMSE of calibration and tuning set) and as well a low tuning set RMSE compared to other values.The premise for this choice was that less overfitting leads to better generalisation capacity of the model.

Once the optimal $\beta = 0.02$ was found, the best kernel and batch sizes were explored. Figure 4 shows the RMSE maps with respect to kernel and batch sizes for the calibration (Figure 4A) and tuning (Figure 4B) sets and for the difference between both (Figure 4C). The black contour lines on the two first maps show the contour level corresponding to 1/3 of the scale, and it was used to empirically define basins of attraction. Since there was no clear one-to-one match between minima in the calibration and tuning maps, we used these criteria to broadly define the basins of attraction or areas of the hyperparameter space where the model performance tends to be more consistent. By overlapping/intersecting these maps (see Supplementary Materials for details) we found 8 points/models that were common to both areas (white dots in the figure) and we pick the 3 (in red) with the lowest tuning-cal RMSE. For the chosen models (2, 3 and 6), their hyperparameters and performance on the original and new test sets were summarised in Table 2.

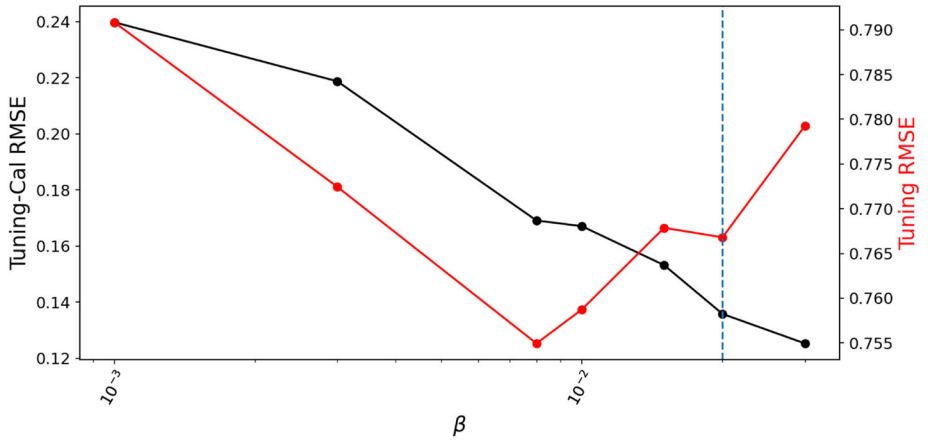

FIGURE 3 Criteria for selecting the strength of the weights regularisation. In black, tuning-cal root-mean-square error (RMSE) and in red tuning RMSE as a function of regularisation parameter β. The vertical blue dashed line represents the points where the best compromise between the cal-tuning RMSE and tuning RMSE was found

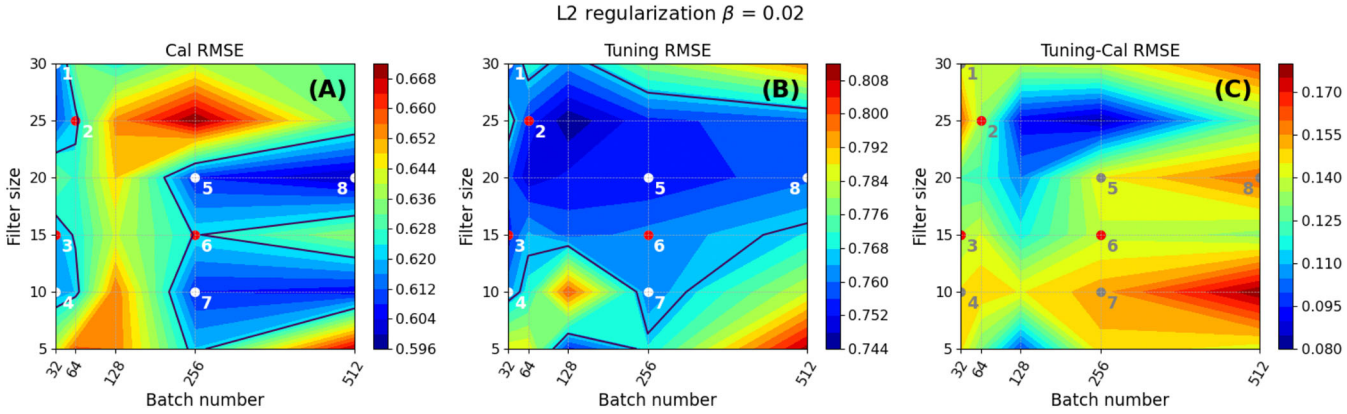

FIGURE 4 The grid search root-mean-square error (RMSE) plots for exploring the effect of batch and filter size. (A) Calibration set, (B) tuning set and (C) RMSE calibration minus RMSE tuning set to judge overfitting. The black contours in (a) and (B) indicate 1/3 of the scale and are used to define basins of attraction in hyperparameter space. From the identified models in these areas, the three models with lowest tuning-cal (marked as red dots) were identified

TABLE 2 The three models found during the hyperparameters optimisation and their performance on the independent test set

Model number	β	Batch size	Kernel size	Test RMSE (%)
2	0.02	64	25	1.053
3	0.02	32	15	0.814
6	0.02	256	15	0.619

For Model 6, the performance was drastically improved by remodelling using data from the year 2018 as the tuning data set. A reason could be that this model was able to reach a model by learning information related to season variability. However, it was unclear what exact information led to such an improvement in the model performance because NIR technology is a non-specific technique where most of the information is captured as highly overlapping spectral responses of underlying physicochemical processes. The RMSE was decreased from 1.053% to 0.619% and was also better than the recalibration of the PLS model, that is, RMSE = 0.68% (Figure 5A). The difference in the RMSEP's of DL and PLS model may not be of high significance considering the high-uncertainty associated with the NIR measurements, however, in a practical use the user usually prefers models with lower errors. A key point to note was that until this stage the reported model accuracies were without any calibration transfer/model adaptation and were

solely based on the open-access mango data measured on the different instrument compared to the one used in this study. In the next section, the effect of a deep calibration transfer on the model performance was presented. From this point onward, due to the best performance, Model 6 was used in the following part of study. The hyperparameters of Model 6 were kernel size = 15, batch size = 256 and $\beta = 0.02$.

3.3 | Deep calibration transfer and PLS recalibration

The open-access mango data set was measured on a different instrument compared to the instrument used in this study. Hence, according to the necessity of chemometrics, a calibration transfer was needed to remove the instrumental differences before model application. Since the primary instrument was not available, traditional standard-based calibration transfer chemometric approaches cannot be used. In this scenario, a possibility could be to develop a new calibration with data from the new instrument. The performance of a new PLS calibration (5 LVs) developed using solely the data of the new instrument was shown in Figure 6C. Additionally, for comparison purposes, another PLS (7 LVs) model was developed by combining the open-access mango data with some data from the new instrument. The performance of the updated PLS was shown in Figure 6D. Finally, the new DL model based on open-access mango data was also updated using some data measured on the new instrument using the TL approach. The TL used in this study consisted in initializing the first five model layers with pre-trained values (of Model 6) and training these layers using 60% of the new test set (fine-tune set). The final layer of the model was trained from scratch (i.e., 'He_Normal' weight initialisation). This procedure allows for a model that has already learned some important patterns in the original data, to evolve to encompass the variability on the new data set, in this study, the new season and new instrument related variability. The final model evaluation was performed by using the retrained model on the remaining 40% of the new test set.

The updating of the model with TL reached the lowest RMSE = 0.518% (Table 3). The new PLS model developed solely on the data from the new instrument showed the second lowest RMSE = 0.598%, while the recalibrated PLS showed the highest RMSE = 0.663% (Table 3). Such a superior performance of the DL model shows that updating the DL model was needed to compensate for the instrument-specific variability and should be performed in practice when a DL model must be used in a new instrument. Although in this current study, we do not have access to the primary Felix F750 equipment used for acquisition of the global mango fruit data set and were not able to show a clear difference between the response of the two Felix 750 instruments. However, the improvement reached after the TL suggests that the TL had to compensate for the instrumental differences. Such instrument differences in NIR spectrometer exist due to several reasons such as the sensitivity of the detector, the light source and many different electric and optic components used in spectrometers.[28]

In the earlier part of this study, the performance of TL was showed using a fine-tune set of 60% of data measured on the new instrument. Those 60% of data were roughly ~300 spectral and reference DM measurements. However, in

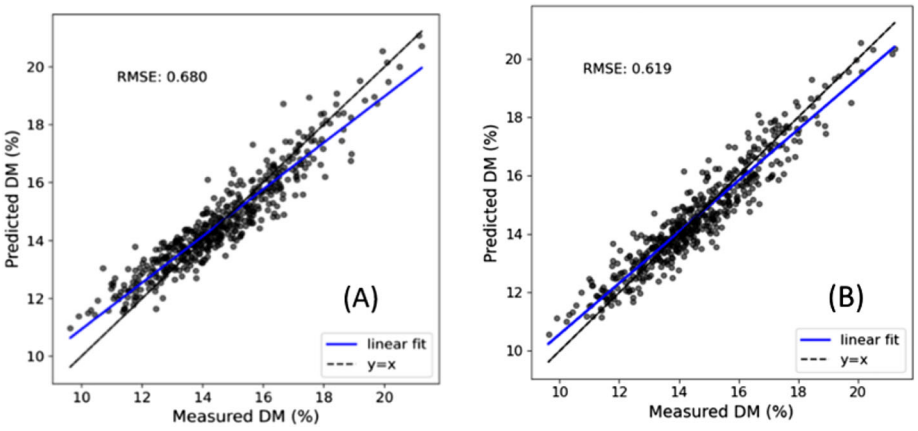

FIGURE 5 (A) Performance of PLS model recalibrated by combining data from calibration and test set of mango data set and tested on the new season harvest. Performance of the new DL model developed by treating test set of mango data set as model validation set and tested on the new season harvest (B) model 6 (see Table 2)

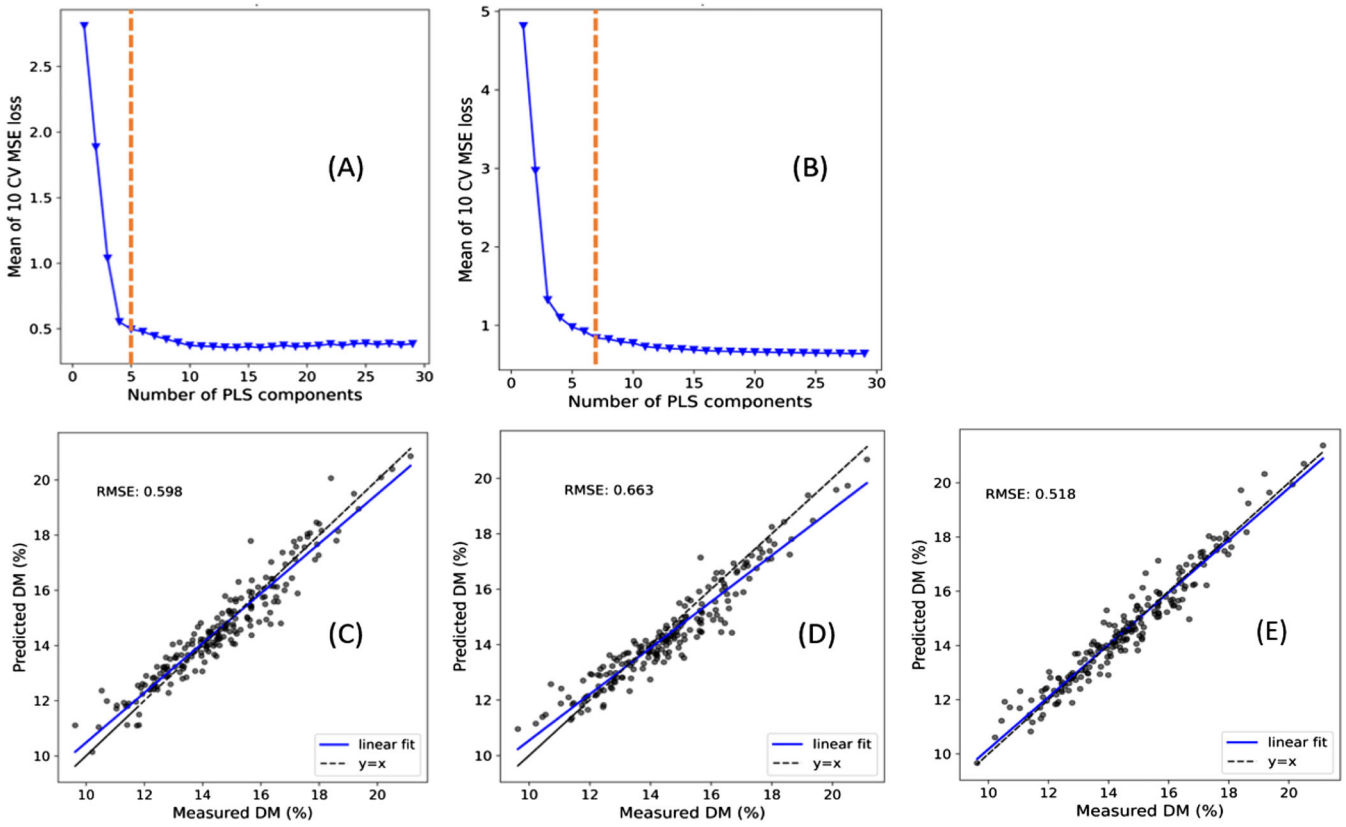

FIGURE 6 (A) Evolution of cross-validation MSE for development of new PLS model based on data from new harvest, five latent variables (LVs) were selected, (B) evolution of cross-validation MSE for development of new PLS model based on combining mango data set and some data from new harvest, seven LVs were selected, (C) PLS model made from new harvest and tested new harvest data, (D) PLS model combining open-access mango data with some data from new harvest and tested on test set from new harvest and (E) fine-tuned DL model on test set from new harvest

TABLE 3 A summary of the performance of DL and PLS models after updating with some new data measured on new instrument

	DL model (RMSEP DM %)	PLS model (RMSEP DM %)
Before transfer learning	0.619	0.754
After transfer learning	0.518	0.663

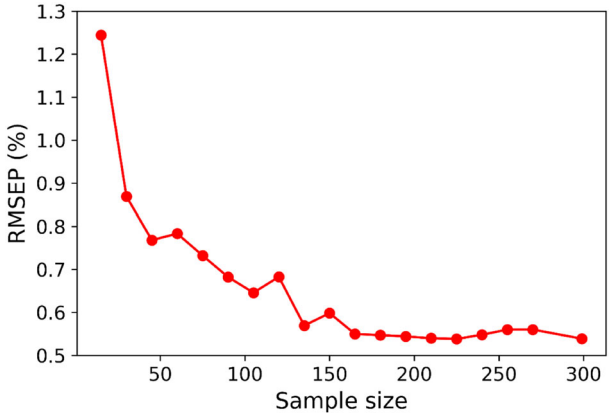

FIGURE 7 A summary of posterior analysis on the effect of samples size on the RMSEP for transfer learning modelling

practice, the user could be interested to measure as minimum samples as possible to make TL a profitable and time-saving task. Therefore, in that regard, a posterior analysis on the effect of the sample size on the performance of the TL model on the independent test set measured on the new instrument was performed. In Figure 7, it can be noted that although in this work nearly 300 spectral measurements were used for the TL modelling. However, the reduction in the RMSEP of TL model stabilised at nearly 150 samples, suggesting that the TL can be performed with a smaller number of samples.

4 | CONCLUSIONS

This study performed independent validation of a global DL mango NIR model[26] with a local experiment on mangoes of a new season and measured with a different instrument. The primary results showed that a direct application of the DL model in the new season and new-instrument data performed poorer compared to its performance reported previously on the test set of open-access mango data set. On the contrary, the PLS model performed better than the DL model when used on data measured with a new instrument. Such an inferior performance of DL model shows that the DL model was much more sensitive to the seasonal/instrument change compared to traditional latent variable based approaches such as PLS. There were two main causes hypothesised for the inferior performance of the DL model. The first was that the instrument used in this study was different compared to what used to measure the data on which the primary DL model was developed. The second reason was that the primary model[26] was not tuned to cover the seasonal variability. In this study, to cover the seasonal variability, a new DL model was developed using the open-access mango data set but using the data from the year 2018 season as the tuning set. Later, the new DL model was transferred to the new instrument in a standard free manner using transfer learning. The DL model reached the lowest RMSEP = 0.518%, which was better than making a new PLS model or recalibrating the old PLS model. Although, the difference in the RMSEP's of DL and PLS model may not be of high scientific significance considering the high-uncertainty associated with the NIR measurements, however, in a practical use, the user usually prefers models with lower errors. The DL modelling can become a practical tool to perform NIR modelling of fresh fruit property. For example, the DL model developed in this study can be used on any similar NIR instrument with a basic step of transfer learning to adjust the model to the local features of the new instrument. DL modelling can increase the scalability and wider usage of NIR models, especially it can support portable spectroscopy where recalibrating each instrument from scratch can be impractical. Future studies will include validating and updating the model in future harvest seasons.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/cem.3367.

DATA AVAILABILITY STATEMENT

Made available on genuine request.

ORCID

Puneet Mishra https://orcid.org/0000-0001-8895-798X
Dário Passos https://orcid.org/0000-0002-5345-5119

REFERENCES

1. Monago-Maraña O, Domínguez-Manzano J, Muñoz de la Peña A, Durán-Merás I. Second-order calibration in combination with fluorescence fibre-optic data modelling as a novel approach for monitoring the maturation stage of plums. *Chemom Intel Lab Syst*. 2020;199:103980.
2. Liu C, Yang SX, Li X, Xu L, Deng L. Noise level penalizing robust Gaussian process regression for NIR spectroscopy quantitative analysis. *Chemom Intel Lab Syst*. 2020;201:104014.
3. Jia B, Wang W, Ni X, et al. Essential processing methods of hyperspectral images of agricultural and food products. *Chemom Intel Lab Syst*. 2020;198:103936.
4. Ye D, Sun L, Tan W, Che W, Yang M. Detecting and classifying minor bruised potato based on hyperspectral imaging. *Chemom Intel Lab Syst*. 2018;177:129-139.
5. Walsh KB, McGlone VA, Han DH. The uses of near infra-red spectroscopy in postharvest decision support: a review. *Postharvest Biol Technol*. 2020;163:111139.

6. Walsh KB, Blasco J, Zude-Sasse M, Sun X. Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: the science behind three decades of commercial use. *Postharvest Biol Technol*. 2020;168:111246.
7. Saeys W, Do Trong NN, Van Beers R, Nicolai BM. *Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: a review, Postharvest Biology and Technology, 158*. Vol. 158; 2019:110981.
8. Nicolai BM, Beullens K, Bobelyn E, et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biol Technol*. 2007;46(2):99-118.
9. Mishra P, Lohumi S, Ahmad Khan H, Nordon A. Close-range hyperspectral imaging of whole plants for digital phenotyping: recent applications and illumination correction approaches. *Comput Electron Agric*. 2020;178:105780.
10. Mishra P, Asaari MSM, Herrero-Langreo A, Lohumi S, Diezma B, Scheunders P. Close range hyperspectral imaging of plants: a review. *Biosyst Eng*. 2017;164:49-67.
11. Mishra P, Woltering E, Brouwer B, Hogeveen-van Echtelt E. Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach. *Postharvest Biol Technol*. 2021;171:111348.
12. Mishra P, Marini F, Brouwer B, et al. Sequential fusion of information from two portable spectrometers for improved prediction of moisture and soluble solids content in pear fruit. *Talanta*. 2021;223:121733.
13. Mazhar M, Joyce D, Lisle A, Collins R, Hofman P. *Comparison of firmness meters for measuring 'Hass' avocado fruit firmness, International Society for Horticultural Science (ISHS)*. Belgium: Leuven; 2016:163-170.
14. Li M, Qian ZQ, Shi BW, Medlicott J, East A. Evaluating the performance of a consumer scale SCiO (TM) molecular sensor to predict quality of horticultural products. *Postharvest Biol Technol*. 2018;145:183-192.
15. Crocombe RA. Portable Spectroscopy. *Appl Spectrosc*. 2018;72(12):1701-1751.
16. Porep JU, Kammerer DR, Carle R. On-line application of near infrared (NIR) spectroscopy in food production. *Trends Food Sci Technol*. 2015;46(2):211-230.
17. Mishra P, Roger JM, Rutledge DN, Woltering E. Two standard-free approaches to correct for external influences on near-infrared spectra to make models widely applicable. *Postharvest Biol Technol*. 2020;170:111326.
18. Mishra P, Roger JM, Marini F, Biancolillo A, Rutledge DN. 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;175:111414.
19. Mishra P, Nikzad-Langerodi R. Partial least square regression versus domain invariant partial least square regression with application to near-infrared spectroscopy of fresh fruit. *Infrared Phys Technol*. 2020;111:103547.
20. Zheng W, Bai Y, Luo H, Li Y, Yang X, Zhang B. Self-adaptive models for predicting soluble solid content of blueberries with biological variability by using near-infrared spectroscopy and chemometrics. *Postharvest Biol Technol*. 2020;169:111286.
21. Bobelyn E, Serban A-S, Nicu M, Lammertyn J, Nicolai BM, Saeys W. Postharvest quality of apple predicted by NIR-spectroscopy: study of the effect of biological variability on spectra and model performance. *Postharvest Biol Technol*. 2010;55(3):133-143.
22. Anderson NT, Walsh KB, Flynn JR, Walsh JP. 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*. 2021;171:111358.
23. Anderson NT, Walsh KB, Subedi PP, Hayes CH. Achieving robustness across season, location and cultivar for a NIRS model for intact mango fruit dry matter content. *Postharvest Biol Technol*. 2020;168:111202.
24. Pasquini C. Near infrared spectroscopy: a mature analytical technique with new perspectives—a review. *Anal Chim Acta*. 2018;1026:8-36.
25. Anderson N, Walsh K, Subedi P, Mango DMC, Spectra Anderson, et al. Mendley. *Mendley Data*. 2020;2020.
26. Mishra P, Passos D. A synergistic use of chemometrics and deep learning improved the predictive performance of near-infrared spectroscopy models for dry matter prediction in mango fruit. *Chemom Intel Lab Syst*. 2021;212:104287.
27. Mishra P, Passos D. Realizing transfer learning for updating deep learning models of spectral data to be used in a new scenario. In: *Chemometrics and Intelligent Laboratory Systems*; 2021:104283.
28. Mishra P, Nikzad-Langerodi R, Marini F, et al. Are standard sample measurements still needed to transfer multivariate calibration models between near-infrared spectrometers? The answer is not always. *TrAC Trends Anal Chem*. 2021;143:116331.
29. Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem*. 1964;36(8):1627-1639.
30. Barnes RJ, Dhanoa MS, Lister SJ. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Appl Spectrosc*. 1989;43(5):772-777.
31. Cui C, Fearn T. Modern practical convolutional neural networks for multivariate regression: applications to NIR calibration. *Chemom Intel Lab Syst*. 2018;182:9-20.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Mishra P, Passos D. Deep chemometrics: Validation and transfer of a global deep near-infrared fruit model to use it on a new portable instrument. *Journal of Chemometrics*. 2021;e3367. https://doi.org/10.1002/cem.3367