Deep learning for near-infrared spectral data modelling: Hypes and benefits

Puneet Mishra, Dário Passos, Federico Marini, Junli Xu, Jose Amigo, Aoife Gowen, Jeroen Jansen, Alessandra Biancolillo, Jean-Michel Roger, Douglas Rutledge, et al.

Trends in Analytical Chemistry 157 (2022) 116804

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

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

Deep learning for near-infrared spectral data modelling: Hypes and benefits

Puneet Mishra [a, *], Dário Passos [b], Federico Marini [c], Junli Xu [d], Jose M. Amigo [e, f], Aoife A. Gowen [d], Jeroen J. Jansen [g], Alessandra Biancolillo [h], Jean Michel Roger [i, j], Douglas N. Rutledge [j, k], Alison Nordon [l]

[a] Wageningen Food and Biobased Research, Bornse Weilanden 9, P.O. Box 17, 6700AA, Wageningen, the Netherlands
[b] CEOT and Physics Department, University of Algarve, Campus de Gambelas, FCT Ed.2, 8005-189, Faro, Portugal
[c] Department of Chemistry, University of Rome "La Sapienza", Piazzale Aldo Moro 5, 00185, Rome, Italy
[d] School of Biosystems and Food Engineering, University College Dublin (UCD), Belfield, Dublin 4, Ireland
[e] IKERBASQUE, Basque Foundation for Science, Bilbao, 48011, Spain
[f] Department of Analytical Chemistry, University of the Basque Country UPV/EHU, P.O. Box 644, Bilbao, Basque Country, 48080, Spain
[g] Radboud University, Institute for Molecules and Materials, P.O. Box 9010, 6500 GL, Nijmegen, the Netherlands
[h] University of L'Aquila, Department of Physical and Chemical Sciences, Via Vetoio, 67100, Coppito, L'Aquila, Italy
[i] ITAP, INRAE Montpellier Institut Agro, University Montpellier, Montpellier, France
[j] ChemHouse Research Group, Montpellier, France
[k] National Wine and Grape Industry Centre, Charles Sturt University, Wagga Wagga, Australia
[l] WestCHEM, Department of Pure and Applied Chemistry and Centre for Process Analytics and Control Technology, University of Strathclyde, Glasgow, G1 1XL, United Kingdom

ARTICLE INFO

Article history:
Received 30 July 2022
Received in revised form
10 September 2022
Accepted 18 October 2022
Available online 21 October 2022

Keywords:
Artificial intelligence
Neural networks
NIR
Near-infrared
Spectroscopy
Chemometrics

ABSTRACT

Deep learning (DL) is emerging as a new tool to model spectral data acquired in analytical experiments. Although applications are flourishing, there is also much interest currently observed in the scientific community on the use of DL for spectral data modelling. This paper provides a critical and comprehensive review of the major benefits, and potential pitfalls, of current DL tecnhiques used for spectral data modelling. Although this work focuses on DL for the modelling of near-infrared (NIR) spectral data in chemometric tasks, many of the findings can be expanded to cover other spectral techniques. Finally, empirical guidelines on the best practice for the use of DL for the modelling of spectral data are provided.

1. Introduction

Chemometrics is a discipline in which artificial intelligence (AI) has had a key role for decades, merging the value of "chemical" and "analytical" intelligence [1,2]. Over the years, artificial neural networks (ANNs) have been used in chemometrics [3] but have failed to demonstrate clear advantages over, e.g., partial least squares (PLS) for classification and prediction. In recent years, however, deep learning (DL) has been able to outperform PLS in several comparative studies on predictive power using NIR spectroscopic data [4–8], sparking interest in the community.

Near infrared (NIR) spectra are manifestations of linear and non-linear combinations of molecular vibrations (combination bands and overtones). These spectra also contain signals resulting from physical effects, such as baselines due to light scattering, multiplicative effects due to pathlength variations and temperature-dependent peak shifts. Classical chemometric approaches (mainly principal component analysis PCA- and PLS-based techniques) combined with knowledge-driven spectroscopic pre-processing have proved to be successful in modelling NIR data. However, to further push model performances with the intention of handling data non-linearities, non-linear methods have been used [9].

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

https://doi.org/10.1016/j.trac.2022.116804

Nevertheless, non-linear methods (e.g., based on kernel support vector machines) may present overfitting problems due to the lack of self-regularization mechanisms (i.e., too many unbounded degrees of freedom) [10,11]. As previously noted, ANNs have been used in the chemometric domain for many years, however, there is a difference between the recently developed deep NNs and the traditional ANNs. ANNs. like most machine learning (ML) algorithms, require pre-extracted features from the spectra as inputs (Fig. 1A), while DL architectures automatically integrate feature extraction (Fig. 1B), including specific proxies for spectroscopic preprocessing [12,13]. This means that DL includes both feature extraction and data-driven preprocessing in model optimization. DL can furthermore incorporate many more layers than ANNs of different types; up to several hundreds of layers with millions of parameters may be trained (which may in turn lead to overfitting). This is possible due to increased computational power, graphics processing units (GPUs), improved regularization techniques and the development of advanced model optimization approaches.

DL for chemometrics is still in its infancy compared to its use in computer vision (CV) and natural language processing (NLP). Several factors may be responsible for this delay in adoption. For example, developments in CV and NLP were possible due to early adoption of DL by commercial technology companies and by the curation of very large open data sets. Meanwhile, in the domain of chemometrics, it is rare to have large open NIR data sets with several thousand samples. Since industrial quality control data are often closed and experimental data are resource-intensive and therefore much sparser than computer vision, they are rarely made open to the scientific community. Also, recent developments in DL are based on the C and Python programming languages, while traditionally, the chemometric tools/resources are mainly available in MATLAB or R.

As with any other experimental profile, techniques for modelling NIR spectral data can be broadly grouped into unsupervised and supervised. DL is also broad in scope and has been tailored to both supervised and unsupervised modelling of NIR data. Examples

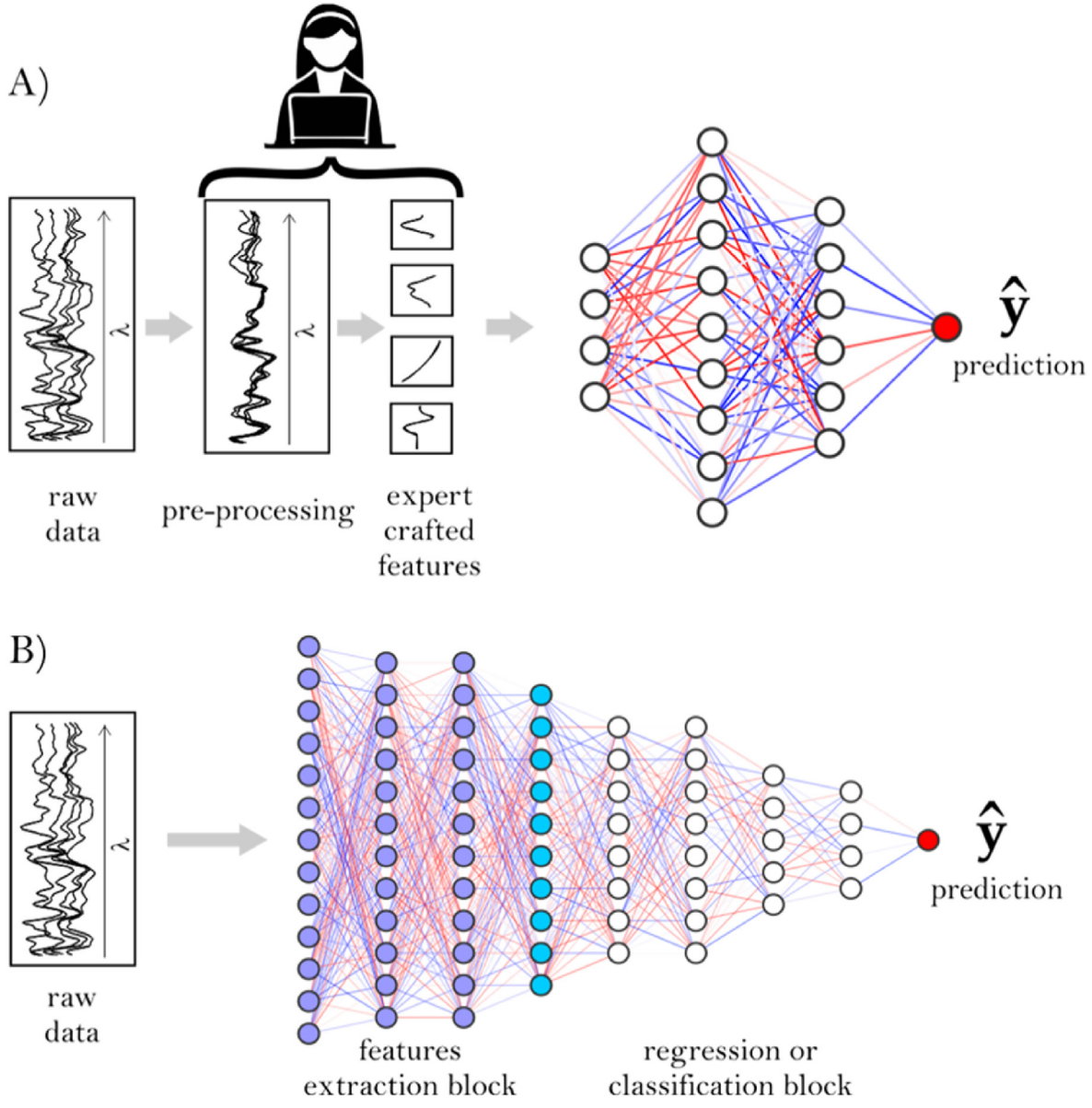

Fig. 1. (A) Classical ANN for data modelling, and (B) DL convolutional neural network (CNN) approach, which includes joint feature extraction and model building.

for supervised modelling are convolutional neural networks (CNNs) [4,14], while for unsupervised modelling, autoencoder (AE)-based methods have been used [15]. DL modelling of spectral data analysis has been used for prediction [16], calibration transfer [17], model updating [18] and spectral image processing [19].

Despite its promising early results, DL methods in spectroscopy have not yet been subjected to systematic applicability studies. The linear (or near-linearizable) nature of the chemical information contained in NIR spectra is often leveraged to produce "simpler" models such as PLS. In recent years, several new chemometric methods have been developed for spectral data modelling. Given these new developments and the early results from DL chemometric applications, DL is often seen as a new tool, considered by some as very promising and by others as just too complex and overrated.

Based on the literature published over the past five years, this work provides a critical and comprehensive review of the potential benefits of DL focused on NIR spectral data modelling, and on which developments in DL may provide chemical value. Finally, clear guidelines on the best practice for the use of DL for the modelling of spectral data are provided and emerging research directions are identified.

2. Open questions in chemometrics for DL-based modelling of NIR data

2.1. Does DL eliminate the need for spectral pre-processing?

Pre-processing is an integral part of the modelling of NIR data. Spectral data can suffer from a wide range of artefacts such as noise, sloping baselines and/or offsets, additive and multiplicative effects, or combinations of these contributions. In a typical chemometric modelling pipeline, either exhaustive search [20] or pre-processing ensembles [21–23] have been proposed to improve models. One of the main drawbacks of these strategies is that they rely on already existing pre-processing approaches. This limitation is especially relevant in the case where datasets are large. Heterogeneous datasets may however include different artefacts from different experimental conditions. Some recent studies [4,12,14,16,24,25] have hinted that DL models can automatically transform data to their most suitable form for predictive modelling. These spectral transformations are done by convolutional layers that are directly connected to other layers in the model. Interestingly [4], has shown empirically the similarity between 1st derivative pre-processed and automatically pre-processed spectra from the convolutional layers of a DL model of their data. Recent works have also shown that DL models can benefit from the use of classical spectral pre-processing approaches [7,12,24–28]. It has been found that, for certain problems, classical pre-processing approaches allowed a faster convergence of DL models [26] and avoided model overfitting [27].

Two recent works [13,29] demonstrated the possibility to integrate pre-processing operations as specialized model layers in DL, which provided better results than those from a single traditional spectral pre-processing approach before either linear or non-linear neural networks [13].

2.2. Can DL be performed on small spectral data sets?

The typical view is that DL models require large amounts of data to train properly. In the area of computer vision, it is common to encounter large image data sets, such as the Aff-Wild2 facial recognition data set (~2,800,000 samples) or the MS COCO data set (2,500,000 samples). While these data size requirements are surely true for the large/very deep models used in CV and NLP, it is not clear if this is the case for the much shallower architectures (up to ~15 layers) used so far for chemometrics tasks. From a technical point of view, through the diligent implementation of regularization techniques, training a DL model of 20 layers or more is possible, even for smaller (n < 1000 where *n* is the number of samples) data sets.

In the domain of NIR spectroscopy, it is uncommon to encounter large open datasets. It is expected that the continued adoption of new spectral technologies (e.g., handheld spectrometers) will contribute to the inversion of this tendency. Large (e.g., number of independent samples > 10^4) open NIR datasets are not common yet in the spectral community and are currently limited to fruit [9,30,31], soil [5,6], and seeds [28,32]. Most of the initial works on the application of DL to NIR spectral analysis have shown the potential of DL on small spectral data sets [4,12,14,16], comprising even fewer than 100 samples in some cases. It seems surprising that DL models, which usually consist of thousands of free parameters, can be trained with only 100 samples; these works implemented "shallow" NNs using 1 to 3 convolutional layers followed by 1–4 dense layers. Hence, this demonstration using a low number of samples cannot be completely neglected but it must be kept in mind that shallow NNs still have hundreds of parameters, which might be difficult to properly train with a low number of data points. Data augmentation techniques can introduce artificial variation in spectral data that help to stabilize the deep NN training. Beyond improving training stability, recent studies [26,33] showed that data augmentation also benefits model performance.

2.3. Is the comparison of DL models with linear models justifiable?

DL produces complex non-linear models, and often, in the spectral data modelling domain, have their performance compared with linear models such as PLS [15,34]. PLS models, if properly optimized in terms of pre-processing and latent variables, are known to be reliable and the recent literature comparing the performance of DL modelling with PLS put a lot of effort into optimizing the DL models while minimal attention was paid to optimizing the PLS models. For example, in a study involving nitrogen prediction in leaves [15], a wide variety of neural architectures were tested while when it came to PLS modelling only one PLS model without any optimization of pre-processing was presented. This calls into question whether a comprehensive comparison of DL with simple PLS approaches has been done until now, and whether such studies really demonstrate that DL is to be preferred for spectral data modelling.

2.4. What about time requirements and explainability of DL models?

The development of linear models with spectral data, such as PLS [35,36], is usually faster compared to DL and reflects the linearity of well-conducted optical spectroscopy. The most time-consuming part of PLS-based approaches is the optimization of pre-processing and the selection of latent variables. However, approaches such as design of experiments [20] and pre-processing ensembles [21,22,37] have been proposed to reduce the time requirements. In comparison to linear approaches, DL model development is time-consuming and resource-intensive. In practice, training DL models and performing the associated optimization tasks on computers equipped just with GPUs can take hours or even days. If DL can overcome the pre-processing and/or variable selection steps, then a suitable trade-off in terms of computation time might be found. One also must take into consideration that much of the time spent around DL crafting is done by exploring different NN architectures. The PLS algorithm is well established but a "general NN architecture" for spectral analysis (if it exists) has not been found yet.

One of the advantages of most classic chemometrics methods is model interpretation, which allows for inference of sample chemistry. Linear models usually allow for direct interpretation of the importance of spectral features (e.g., through scores and loadings in PLS) to identify key physicochemical components. This type of interpretation is not directly accessible from DL models. DL can accurately predict properties of many spectra, but their arguably "black-box" nature makes it difficult to identify what spectroscopic patterns direct the classification. Currently, the general perception is that DL-based approaches offer little in terms of causality insights and inverse problem responses. However, model interpretability is currently one of the most important topics of research in DL. In NIR analysis, some works have demonstrated the use of class activation mapping (CAM) [38] and the application of perturbation theory to compute regression coefficients [4] to interpret DL models.

3. Potential benefits of DL for spectral data modelling

3.1. DL on large datasets can outperform non-DL chemometric modelling approaches

It has been generally shown that the performance of DL models scales with training data size better than most non-DL chemometric approaches. This has been shown by Ref. [5] for the prediction of soil properties (i.e. organic carbon content, cation exchange capacity and texture) using NIR spectral data. While the authors of [5,39] pointed out that they found no advantage of using DL models with small ($n < 1000$) data sets when compared to PLS and cubist models, they saw clear gains with increasing data size.

In another example, DL was applied to NIR spectra for dry matter prediction of mango fruit [7] with ~10^4 samples; the authors achieved better performance in terms of a lower root mean square error in prediction (RMSEP) than with a wide range of non-DL chemometric approaches [9,30]. Another work on soil spectroscopy [6], which also explored DL approaches based on ~10^4 spectra and values of the corresponding reference properties, achieved better performance than non-DL chemometric approaches. This performance advantage of DL models (ranging from a few to 10% improvement) has also been reported in other works that use NIR spectral data sets with > 10^4 samples [28,32,40]. Hence, based on the current trend, it is justifiable to say that DL for NIR spectral data has real benefits when many samples are available. One should note that the amount of data is not the sole reason for the improved performance of the DL model and that the quality of data acquired is also important. However, in different areas of science such as soil analysis and fruit science (where NIR is extensively used, and large data sets are generated) it is still challenging to obtain fully representative data; usually in such domains data is continuously collected for multiple locations, years, batches etc. to incorporate large variations into the dataset.

3.2. DL can efficiently handle complexities in response variables

One of the known key benefits of DL is its ability to efficiently handle complex response variables, for example, a multi-class classification [32,40] or multi-response predictive model [8,41]. In a recent study [40] and as shown in Fig. 2, as the complexity of the response variable (such as the total number of classes) increased, the DL model could outperform classical chemometric approaches; as the number of classes increased, the performance of the non-DL methods decreased while the DL model maintained its performance. Note that the results in Fig. 2 are based on a single data set and such a comparison on a different data set might yield a different trend. In another recent study for pectin strength prediction [8], a DL model based on data combined from five different pectin formulations performed better than PLS2. The capability of DL to better handle complex responses and particularly multi-response scenarios is probably related to its high plasticity and its improved non-linear pattern recognition capabilities. DL models can better capture second-order correlations between input variables due to the weight-sharing [42] properties of NNs and may extend the range of informative variables that can contribute to the prediction.

3.3. DL facilitates the fusion of data from different modalities

As mentioned previously, one of the merits of DL is its ability to perform automatic feature extraction from the input data. Such a feature extraction is possible through specialized layers such as convolutional layers; for example, a 1-D convolutional operation for a spectral signal, a 2-D convolutional operation for an image, and a 3-D convolution for data cubes or higher order data. From the DL perspective, it is straightforward to apply different convolution operations to different modalities of data and perform data fusion. Currently there are only a few works that have touched on this topic as applied to chemometric tasks [43,44]. In Ref. [46], the authors explained and demonstrated the capability of parallel convolutional based DL models for fusing information from two spectral ranges, i.e., visible and NIR, and demonstrated that the model based on fused information performed slightly better (2% lower RMSEP) than models based on individual spectral ranges. Another study provided a general framework for deep multi-block modelling [43], and used parallel input convolutional/autoencoder networks (Fig. 3). Applications of such deep multiblock approaches are still in their infancy.

3.4. Model sharing and transfer with transfer learning

Accurate and reliable spectral models are of high value due to the associated cost of the reference analyses that are usually required to perform the calibration. In the domain of chemometrics, the transfer and sharing of spectral models is very common [45]. Just like classical chemometric calibration transfer approaches, DL models are also highly flexible for sharing and updating, enabling their use in a new situation using the concept of transfer learning (TL). The key motivation behind TL is that DL models are trained on large datasets, which requires both time and resources, and so it does not make sense to start to train new models from scratch. Hence, the concept of TL allows updating of an existing model for new scenarios by retraining only limited parts of the model such as dense layers or feature extraction layers (Fig. 4). Although TL in spectral data modelling is still a very new topic, recent works have shown that TL is capable of handling multiple tasks such as fine-tuning a DL model on small datasets using a model pre-trained on large data sets [46], updating models to incorporate new variabilities [18], and one of the most common chemometric tasks, i.e., calibration transfer [17,47].

3.5. Joint modelling of spatial and spectral information for spectral imaging

Apart from the modelling of spectra obtained from a point spectrometer, DL is also suitable for modelling spectral imaging datasets [48,49]. As a 3-D data array with two spatial dimensions and one spectral dimension, a spectral image contains rich information that is useful for diverse applications. Currently, in the chemometrics domain the most practiced approach to spectral image processing is to treat the pixels as point spectrometer data and usually limited information about image context is exploited [15,34]. Within the image context, it is beneficial if the model can

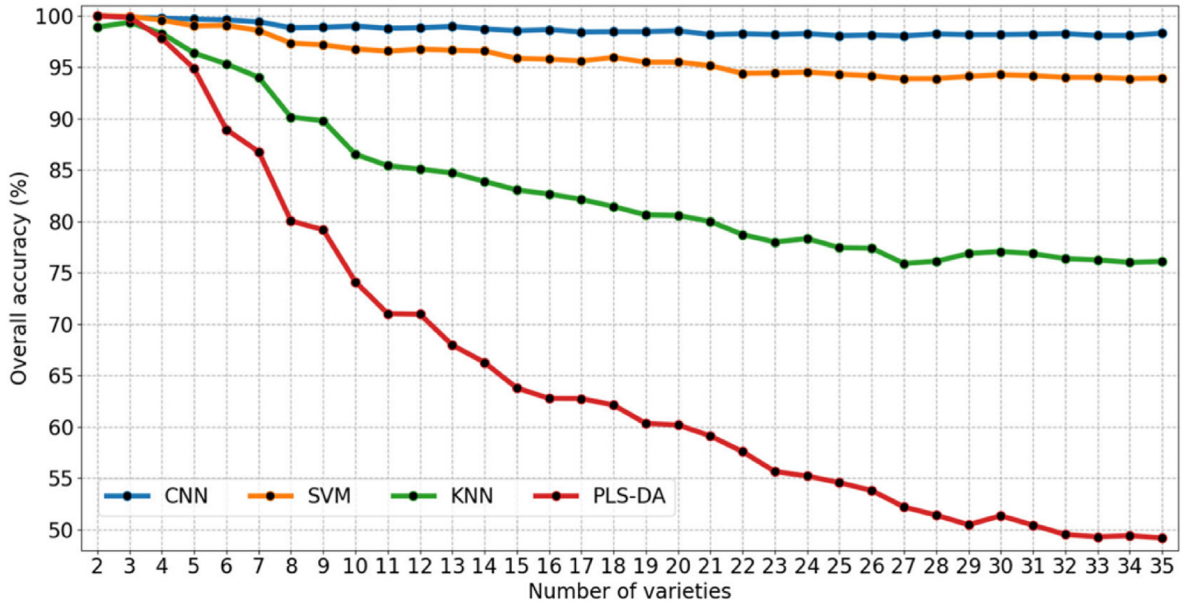

Fig. 2. Performance of different chemometric and DL approaches as a function of the number of classes for classification modelling [40].

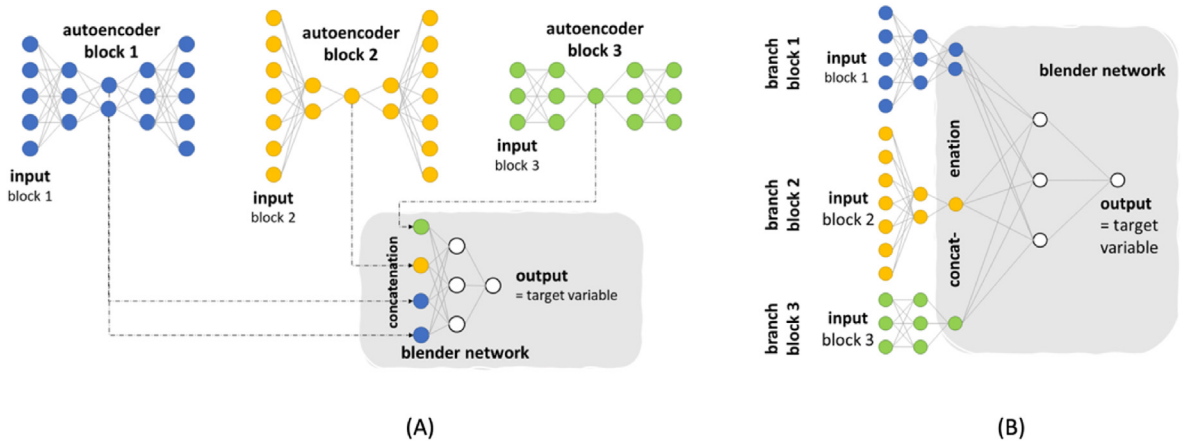

Fig. 3. A graphical summary of deep multiblock modelling approaches [43]. (A) Features are extracted using autoencoders and bottle neck layers are connected to dense layers, and (B) parallel convolutional operations are carried out for data from different modalities and later features are concatenated to build a single model.

take the "spatial coherence" into account because neighboring pixels usually share the same labels/chemical properties unless the imaged scene is highly heterogenous. One of the main benefits of performing DL on spectral images is that developments from the computer vision domain can be directly translated to spectral image processing, as spectral images can be considered as images with more than three color channels. Hence, DL is a powerful tool to process spectral images by allowing modelling of both the spatial and spectral information using 2D or 3D [50] convolutional operations. Performing this type of modelling can allow the combination of contextual information with the chemical information present in the spectroscopy domain [19,27,51]. Several applications of DL for spectral imaging can be found, ranging from semantic segmentation to classification and regression modelling [19,27,51–53].

4. General guidelines for deep spectral data modelling

Some of the following points are general chemometrics recommendations that also apply to DL, and some are specific tips for a guided exploration of DL for spectral data modelling:

Start simple: If the user is trying to solve a chemometric task that involves small datasets (number of independent samples <1000), and that is the main scientific goal, it is best to start with the classical linear models. If the answer is unsatisfactory and a large dataset is available explore DL approaches. Start with shallow DL architectures and increment complexity progressively.

Data set size: The performance of DL tends to improve as a function of data set size. Although shallow DL architectures can be trained with a small number of samples, more complex/deeper models will only reach full potential if sufficient data is available to train all the free parameters in the model. Whenever possible increase the data set, making sure that the additional samples are independent, yet of similar analytical quality and quantitatively comparable to those already present in the data.

Implement data augmentation: Data augmentation techniques like adding perturbed versions of existing spectra (e.g., with extra noise, trends) or additional variables to the input variable space (e.g., concatenation of pre-processing) can help to improve the training stability of larger DL models.

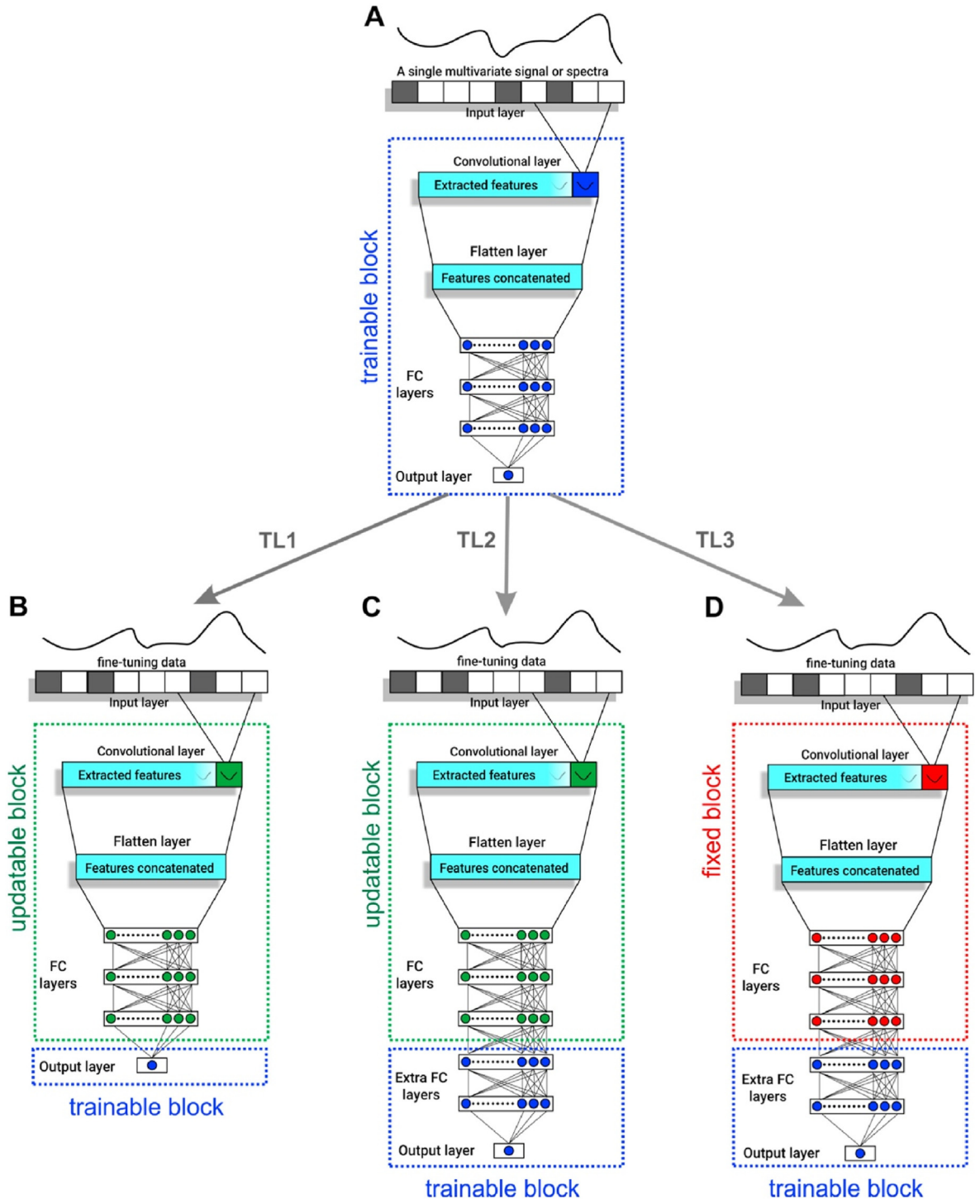

Fig. 4. A summary of deep learning (DL) and transfer learning (TL) architecture for model training and transfer [18]. (A) The DL architecture for primary model training, (B) the model used in TL1, (C) the model used in TL2 and (D) the model used in TL3. Blue represents units that were allowed normal training; green represents units initialized with parameters from the pre-computed base model and are updated using fine-tuning data; red represents units with fixed ("untrainable") parameters loaded from the pre-computed base model.

Model weight sharing and transfer learning: If the number of samples in the dataset is low, then the user can explore model sharing and transfer learning approaches to learn either from the weights of a model trained using large data sets or from updating an existing similar model.

Testing of models on independent batches: As for all types of models, the recommendation is that DL models should be validated using independent batches of data. It is well known in the spectral data modelling community that models are highly sensitive to subtle changes in measurement conditions, such as temperature, instrument, light source, and batch effects. An ideal generalized model should perform well when tested on new batches of data. Also note that traditionally model performance is judged based on parameters such as coefficient of determination or prediction

errors, however, additional analytical figures of merit such as generalized analytical sensitivity [54] have been proposed to judge DL models irrespective of their architecture. Hence, DL practitioners should include parameters such as generalized analytical sensitivity when comparing their models.

Model optimization and hyperparameters tuning: Hyperparameter optimization should be standard practice during the development of a new DL model or architecture; furthermore, justification of the usage of different layers is well appreciated and will help the community to understand the role of different model layers for their practical implementation.

Model interpretability: It is advised that future practitioners should also report the interpretation of models with approaches such as class activation mapping [37] or similar techniques [3].

Comparison with state-of-the-art chemometric approaches: When the objective is to obtain the best result for some specific chemometric task, DL models should be compared with the most recent tools for chemometric modelling of NIR spectral data. If the aim is to explore DL architectures for chemometric tasks then a comparison with classical chemometric approaches is unnecessary. However, in this case, the user should use benchmark approaches to show that the different DL architectures are appropriate for different data sets.

Open data sets: One of the basic foundations of DL is large data sets. It is advocated that, whenever possible, research groups share their datasets with the scientific community using institutional websites or online platforms. As an example, a recent review proposes the use of two data sets for model benchmark purposes. One data set is related to regression analysis and was originally published in Ref. [7], while the other is related to multi-class classification and was made available in Refs. [28,55]. A copy of these datasets is available for download at:

https://github.com/dario-passos/DeepLearning_for_VIS-NIR_Spectra/tree/master/notebooks/Tutorial_on_DL_optimization/datasets/

Open codes: DL is a relatively new topic and many people working in the chemometrics domain are not well acquainted with the programming of DL. Hence, it is very important that people practicing DL should share the code used in their published works so that the scientific community can learn and take developments forward. Some examples of existing code that is currently available are the spectral image processing tutorial [53] and the automated optimization of DL models for spectral classification and regression [55].

Proper indication of computational requirements: In DL models, training and optimization are computationally expensive and time-consuming; when reporting the performance of DL models, it is worthwhile to report the time required for model training along with information about the hardware of the computing systems employed such as GPUs etc.

5. Conclusions and perspectives

DL for spectral data analysis is a very recent topic in chemometrics and can be traced back to just the past five years. Although DL approaches have been used for a wide range of tasks, such as regression, classification, model updating, calibration transfer and spectral image processing, our review reveals that many published studies to date have been performed using a small number of samples and under-optimized models, thus not allowing for generalized conclusions to be made in terms of improvements in model performance as compared to classical chemometric techniques. Several applications of DL models to NIR spectral data analysis have been published, and these are generally targeted towards the prediction of a constituent within a sample, or classification of samples according to variations in composition. However, there are very few studies concerning DL model behavior for spectral data, e.g., how each type of layer impacts the model's dynamics, or the impact of DL architecture on the results. This presents an opportunity for the chemometrics community to expand research into the development of new architectures appropriate for the analysis of NIR spectral data, instead of using architectures developed to tackle problems in other areas. Interesting DL properties found in other research areas, such as CNNs' "spatial invariance", i.e., the capacity for a CNN model to identify an object in an image independently of its position or orientation, raises the question of whether something similar could be achieved in the case of spectra, and how this would translate in terms of model response. Many other interesting DL properties like this example, i.e., observed in the case of other types of data, might be relevant to solve problems in chemometrics. The inclusion of expert knowledge into the DL model in the form of specialized layers for spectral analysis is also another interesting approach. In the limited number of studies that have used large datasets, it has been demonstrated that DL can indeed outperform the traditional chemometrics approaches on various occasions. This reinforces the point that for deep spectral modelling, large spectral data sets with wide variability are key to training more complex, accurate and robust models. Data augmentation for performing DL on smaller data sets and its impact on model generalizability to new batches of data is still a developing area of research where target studies, using well curated datasets could take advantage of the chemometrician's expert knowledge.

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.

Data availability

No data was used for the research described in the article.

Acknowledgments

Junli Xu and Aoife Gowen acknowledge funding from Science Foundation Ireland (SFI) under the investigators programme Proposal ID 15/IA/2984-HyperMicroMacro. Dário Passos acknowledges FCT - Fundação para a Ciência e a Tecnologia, Portugal, for funding CEOT project UIDB/00631/2020 CEOT BASE and UIDP/00631/2020 CEOT PROGRAMÁTICO.

References

[1] F. Marini, Artificial neural networks in foodstuff analyses: trends and perspectives A review, Anal. Chim. Acta 635 (2) (2009) 121–131.
[2] F. Marini, et al., Artificial neural networks in chemometrics: history, examples and perspectives, Microchem. J. 88 (2) (2008) 178–185.
[3] B. Debus, et al., Deep learning in analytical chemistry, TrAC, Trends Anal. Chem. 145 (2021), 116459.
[4] C. Cui, T. Fearn, Modern practical convolutional neural networks for multivariate regression: applications to NIR calibration, Chemometr. Intell. Lab. Syst. 182 (2018) 9–20.
[5] W. Ng, et al., The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data, Soil 6 (2) (2020) 565–578.
[6] W. Ng, et al., Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra, Geoderma 352 (2019) 251–267.
[7] 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, Chemometr. Intell. Lab. Syst., 2021, 104287.

[8] K.A. Einarson, et al., Predicting pectin performance strength using near-infrared spectroscopic data: a comparative evaluation of 1-D convolutional neural network, partial least squares, and ridge regression modeling, J. Chemometr. (2021), e3348.
[9] N.T. Anderson, et al., 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.
[10] J.R.M. Smits, et al., Using artificial neural networks for solving chemical problems: Part I. Multi-layer feed-forward networks, Chemometr. Intell. Lab. Syst. 22 (2) (1994) 165–189.
[11] W.J. Melssen, et al., Using artificial neural networks for solving chemical problems: Part II. Kohonen self-organising feature maps and Hopfield networks, Chemometr. Intell. Lab. Syst. 23 (2) (1994) 267–291.
[12] J. Acquarelli, et al., Convolutional neural networks for vibrational spectroscopic data analysis, Anal. Chim. Acta 954 (2017) 22–31.
[13] R. Helin, et al., On the possible benefits of deep learning for spectral preprocessing, J. Chemometr. (2021), e3374.
[14] S. Malek, F. Melgani, Y. Bazi, One-dimensional convolutional neural networks for spectroscopic signal regression, J. Chemometr. 32 (5) (2018) e2977.
[15] 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.
[16] X. Zhang, et al., DeepSpectra: an end-to-end deep learning approach for quantitative spectral analysis, Anal. Chim. Acta 1058 (2019) 48–57.
[17] P. Mishra, D. Passos, Deep chemometrics: validation and transfer of a global deep near-infrared fruit model to use it on a new portable instrument, J. Chemometr. (2021), e3367.
[18] P. Mishra, D. Passos, Realizing Transfer Learning for Updating Deep Learning Models of Spectral Data to Be Used in a New Scenario, Chemometrics and Intelligent Laboratory Systems, 2021, 104283.
[19] P. Mishra, I. Herrmann, GAN Meets Chemometrics: Segmenting Spectral Images with Pixel2pixel Image Translation with Conditional Generative Adversarial Networks 215, Chemometr. Intell. Lab. Syst., 2021, 104362.
[20] J. Gerretzen, et al., Simple and effective way for data preprocessing selection based on design of experiments, Anal. Chem. 87 (24) (2015) 12096–12103.
[21] J.-M. Roger, A. Biancolillo, F. Marini, Sequential preprocessing through ORThogonalization (SPORT) and its application to near infrared spectroscopy, Chemometr. Intell. Lab. Syst. 199 (2020), 103975.
[22] P. Mishra, et al., Parallel Pre-processing through Orthogonalization (PORTO) and its Application to Near-Infrared Spectroscopy, Chemometr. Intell. Lab. Syst., 2020, 104190.
[23] L. Xu, et al., Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration, Anal. Chim. Acta 616 (2) (2008) 138–143.
[24] X. Wang, et al., End-to-end analysis modeling of vibrational spectroscopy based on deep learning approach, J. Chemometr. 34 (10) (2020) e3291.
[25] Y.-Y. Chen, Z.-B. Wang, End-to-end quantitative analysis modeling of near-infrared spectroscopy based on convolutional neural network, J. Chemometr. 33 (5) (2019) e3122.
[26] U. Blazhko, et al., Comparison of augmentation and pre-processing for deep learning and chemometric classification of infrared spectra, Chemometr. Intell. Lab. Syst. 215 (2021), 104367.
[27] P. Mishra, et al., Complementary chemometrics and deep learning for semantic segmentation of tall and wide visible and near-infrared spectral images of plants, Comput. Electron. Agric. 186 (2021), 106226.
[28] D. Passos, P. Mishra, An automated deep learning pipeline based on advanced optimisations for leveraging spectral classification modelling, Chemometr. Intell. Lab. Syst. 215 (2021), 104354.
[29] J. Dong, et al., A practical convolutional neural network model for discriminating Raman spectra of human and animal blood, J. Chemometr. 33 (11) (2019) e3184.
[30] N.T. Anderson, et al., Achieving robustness across season, location and cultivar for a NIRS model for intact mango fruit dry matter content, Postharvest Biol. Technol. 168 (2020), 111202.
[31] N. Anderson, K. Walsh, P. Subedi, Mango DMC and spectra Anderson et al, vol. 2020, 2020. Mendley: Mendley data.
[32] L. Zhou, et al., Wheat kernel variety identification based on a large near-infrared spectral dataset and a Novel deep learning-based feature selection method, Front. Plant Sci. 11 (2020) 1682.
[33] E.J. Bjerrum, M. Glahder, T. Skov, Data Augmentation of Spectral Data for Convolutional Neural Network (CNN) Based Deep Chemometrics, 2017. arXiv preprint arXiv:1710.01927.
[34] Z. Xin, et al., A Deep Learning Based Regression Method on Hyperspectral Data for Rapid Prediction of Cadmium Residue in Lettuce Leaves, vol. 200, Chemometrics and Intelligent Laboratory Systems, 2020, 103996.
[35] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometr. Intell. Lab. Syst. 58 (2) (2001) 109–130.
[36] H. Martens, T. Næs. Multivariate Calibration, John Wiley and Sons, New York, 1991.
[37] P. Mishra, et al., New data preprocessing trends based on ensemble of multiple preprocessing techniques, TrAC, Trends Anal. Chem. 132 (2020), 116045.
[38] R.R. Selvaraju, et al., Grad-CAM: visual explanations from deep networks via gradient-based localization, Int. J. Comput. Vis. 128 (2) (2020) 336–359.
[39] J. Padarian, B. Minasny, A.B. McBratney, Using deep learning to predict soil properties from regional spectral data, Geoderma Regional 16 (2019), e00198.
[40] T. Singh, N.M. Garg, S.R.S. Iyengar, Nondestructive identification of barley seeds variety using near-infrared hyperspectral imaging coupled with convolutional neural network, J. Food Process. Eng. (2021), e13821.
[41] P. Mishra, D. Passos, Multi-output 1-dimensional convolutional neural networks for simultaneous prediction of different traits of fruit based on near-infrared spectroscopy, Postharvest Biol. Technol. 183 (2022), 111741.
[42] J.S. Larsen, L. Clemmensen, Weight sharing and deep learning for spectral data, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP), 2020.
[43] A. Jenul, et al., Multiblock-Networks: A Neural Network Analog to Component Based Methods for Multi-Source Data, 2021. arXiv preprint arXiv:2109.10279.
[44] P. Mishra, D. Passos, Deep multiblock predictive modelling using parallel input convolutional neural networks, Anal. Chim. Acta (2021), 338520.
[45] P. Mishra, 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. 143 (2021), 116331.
[46] J.S. Larsen, L. Clemmensen, Deep Learning for Chemometric and Non-translational Data, 2019. arXiv preprint arXiv:1910.00391.
[47] P. Mishra, D. Passos, Deep calibration transfer: transferring deep learning models between infrared spectroscopy instruments, Infrared Phys. Technol. 117 (2021), 103863.
[48] M.E. Paoletti, et al., Deep learning classifiers for hyperspectral imaging: a review, ISPRS J. Photogrammetry Remote Sens. 158 (2019) 279–317.
[49] C. Wang, et al., A review of deep learning used in the hyperspectral image analysis for agriculture, Artif. Intell. Rev. 54 (7) (2021) 5205–5253.
[50] J. Acquarelli, E. Marchiori, L.M.C. Buydens, T. Tran, T. Van Laarhoven, Spectral-Spatial Classification of Hyperspectral Images: Three Tricks and a New Learning Setting, Remote Sens. 10 (2018) 1156. https://doi.org/10.3390/rs10071156.
[51] P. Mishra, et al., A generic workflow combining deep learning and chemometrics for processing close-range spectral images to detect drought stress in Arabidopsis thaliana to support digital phenotyping, Chemometr. Intell. Lab. Syst. 216 (2021), 104373.
[52] P. Mishra, Deep generative neural networks for spectral image processing, Anal. Chim. Acta (2021), 339308.
[53] J.-L. Xu, et al., Deep learning classifiers for near infrared spectral imaging: a tutorial, J. Spectr. Imaging 9 (2020), a19.
[54] K. Shariat, et al., Sensitivity and generalized analytical sensitivity expressions for quantitative analysis using convolutional neural networks, Anal. Chim. Acta 1192 (2022), 338697.
[55] D. Passos, P. Mishra, A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks, Chemometr. Intell. Lab. Syst. 223 (2022), 104520.