Chemometrics and Intelligent Laboratory Systems 272 (2026) 105664

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Barlow Twins for semi-supervised learning in NIR spectroscopy

Mark Wohlers [a,b,*], Andrew McGlone [a], Eibe Frank [b], Geoffrey Holmes [b]

[a] New Zealand Institute for Bioeconomy Science Limited, Auckland, New Zealand
[b] Department of Computer Science, University of Waikato, Hamilton, New Zealand

ARTICLE INFO

Dataset link: https://github.com/mwohlers/BarlowTwins_NIR

Keywords:
Near-infrared spectroscopy
Semi-supervised learning
Self-supervised learning
Partial least squares regression
Barlow Twins

ABSTRACT

Near-infrared (NIR) spectroscopy is a widely used technology in the horticulture industry for non-destructive fruit grading. Partial Least Squares (PLS) regression is the dominant method for producing fruit quality predictions from measured spectra. Alternative deep learning methods have shown promise, but often require large amounts of labelled data to train. This study proposes a semi-supervised method based on Barlow Twins to include unlabelled data in the training process. We adopt the Barlow Twins method by using repeated measurements on the same fruit from different devices as different "views" to encode into the same latent space and combine the encoder network with a regression head for prediction. Our approach demonstrates improved performance over PLS with up to 17% lower RMSE, especially when the labelled data is limited. The Barlow loss function also improves calibration transfer results.

1. Introduction

Partial Least Squares Regression (PLSR) is a popular and widely adopted method in NIR spectroscopy. While deep learning approaches, particularly convolutional neural networks [1–4] have shown promise, much of the research focuses on architectures and training methodologies for image-based data. In recent years, self-supervised learning methods (SSL) have enabled learning from unlabelled data, but these are underexplored for NIR data. Fortunately, standard deep learning frameworks such as PyTorch [5] and TensorFlow [6] allow for simple specification of network architectures suitable for NIR data and custom loss functions, including weighted combinations of losses, that are applied in SSL methods. This paper investigates the application of the Barlow Twins SSL approach to NIR data. This approach applies a nonstandard loss function that consists of a weighted combination of terms based on the diagonal and off-diagonal components of the cross-correlation matrix between two embedded views of the same sample. By further combining the Barlow Twins loss with a regression mean square error (MSE) loss, we can train models that benefit from the information contained in both labelled data (spectra with reference values) and unlabelled data (spectra without reference measurements) within a semi-supervised framework.

2. Background

2.1. NIR spectroscopy

Near-Infrared spectroscopy provides a nondestructive method to predict the chemical makeup of samples based on the absorbances of different wavelengths in the near-infrared range. It has been applied to many industries, including horticulture, for the prediction of fruit quality measures associated with consumer responses such as dry matter content (DMC) and soluble solids content (SSC).

2.2. Partial least squares

Partial least squares regression has been widely applied in the NIR spectroscopy space. Despite a large number of competing methods, including convolutional networks, PLS remains the dominant method for NIR regression and classification tasks [7]. Due to its popularity and robustness, PLS is often used as a baseline method for comparisons of alternative modelling techniques.

Developed in 1975 by Herman Wold and later modified by Svante Wold and Harald Martens [8], PLS regression addresses the problem of multicollinearity, which is present in NIR data, where responses at different wavelengths are highly correlated. It does this by finding linear combinations of the input variables (wavelengths) to form latent

* Corresponding author at: New Zealand Institute for Bioeconomy Science Limited, Auckland, New Zealand.
E-mail address: mark.wohlers@plantandfood.co.nz (M. Wohlers).

https://doi.org/10.1016/j.chemolab.2026.105664
Received 23 October 2025; Received in revised form 7 February 2026; Accepted 10 February 2026
Available online 14 February 2026

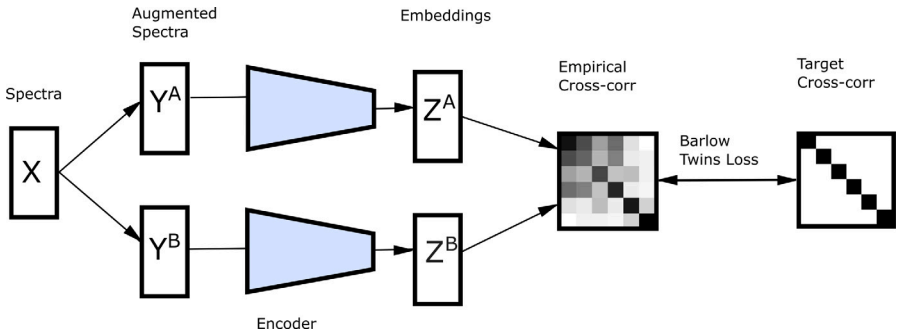

Fig. 1. Barlow Twins loss measures the cross-correlation of embeddings from two related inputs and penalises for how different it is from the identity matrix.

variables. These latent variables are fitted sequentially to maximise the covariance of the latent variable to the target variables (Y), subject to the constraint that the weight vectors have unit norm and successive latent variables are orthogonal. After each latent variable is determined, the X (wavelength) and Y (target) matrices are deflated before finding the next component.

2.3. Self-supervised learning

Self-supervised learning is a form of unsupervised learning that derives discriminatory features from unlabelled data [9]. Examples include reconstructing images from mixed-up jigsaw pieces, which helps learn spatial features [9]. There has been little work on self-supervised learning applied to NIR spectroscopy. Zhang et al. [10] pretrained a dual-branch autoencoder on unlabelled VIS (450–780 nm range) and NIR (783–1500 nm) datasets of mango fruit. The trained encoder was then used to encode smaller labelled datasets for regression tasks. They found that their method achieved results within 99% of the best results using only 10% of the labelled data.

2.4. Barlow Twins

Barlow Twins [11] is a self-supervised learning method originally applied to image data. The method involves generating pairs of augmented views of a given image, which are then fed into an encoder. This encoder is trained so that the values of any given latent variable in the encoder's output are highly correlated between the two augmented pairs, while the values of different latent variables have low correlation. The loss function $\mathcal{L}_{BT}$ is calculated from the cross-correlation matrix of the two latent samples. To minimise the loss, the off-diagonal elements should tend to 0, and the diagonal elements to 1, approximating the identity matrix (Fig. 1). The method has the advantage of providing latent variables that avoid redundancy and are invariant to the distortions used in the augmentation, while preventing collaps to constant outputs.

The Barlow Twins loss function is defined as a weighted sum of the invariance and redundancy reduction terms. The λ parameter controls this weighting :

$$\mathcal{L}_{BT} \triangleq \underbrace{\sum_i (1 - C_{ii})^2}_{\text{invariance term}} + \lambda \underbrace{\sum_i \sum_{j \neq i} C_{ij}^2}_{\substack{\text{redundancy} \\ \text{reduction term}}} \tag{1}$$

The cross-correlation matrix elements C_{ij} are calculated as:

$$C_{ij} \triangleq \frac{\sum_b z^A_{b,i} z^B_{b,j}}{\sqrt{\sum_b (z^A_{b,i})^2} \sqrt{\sum_b (z^B_{b,j})^2}} \tag{2}$$

In the current study, we treat spectra from different devices measuring the same fruit as different views of the same object, similar to capturing the object with two different cameras.

2.5. Semi-supervised learning

Semi-supervised learning is similar to self-supervised learning in that it derives information from unlabelled data [12]. The distinction is that it does this in combination with a (often smaller) labelled dataset. In a deep learning setting, unlabelled data could be used to train an encoder with a Barlow loss function, while a regression head is trained on the labelled set with a mean squared error (MSE) loss function. Several recent publications have examined the potential of applying semi-supervised learning to NIR spectroscopy [13,14]. Said et al. [13] looked at semi-supervised learning for predicting milk fat content. Their method used autoencoders (AE) trained on unlabelled data, while training a regression head connected to the encoder on a smaller subset of labelled data. The autoencoder and regression head were trained simultaneously, with the average of the autoencoder reconstruction MSE and the regression MSE as the loss function. They found improved results compared to the same deep learning regression model trained on the labelled data only, including a reduction in root mean square error (RMSE) from 0.287 to 0.221 with only 35% of the labelled and 65% unlabelled data. Mishra and Woltering [14] did not incorporate unlabelled data into their method and instead focused on robust regression by down-weighting outliers.

3. Proposed framework

3.1. Barlow Twins for spectral data

We propose using the Barlow Twins loss to train an encoder model combined with a regression head that uses MSE loss. To adapt the Barlow method to NIR spectra, suitable augmentation methods are needed. Recently, Dhaini et al. [15] examined contrastive learning for hyperspectral image classification using various augmentation techniques, including spectral shift, spectral flipping, and scattering using Hapke's model [16]. Spectra could also be augmented by adding random multivariate normal samples as in [17]. Alternatively, when samples are measured by multiple devices, pairs of measurements of the same sample by different devices can be used as different views (see Fig. 2).

3.2. Links to PLS

A PLS regression model can be approximated using this framework by employing an appropriately configured neural network. The encoder network should consist of a single linear dense layer with the number of neurons equal to the number of latent variables, and with a unit norm constraint on the weights. To train a PLS-type regression type model, no augmentation is used; that is, the pairs of spectra are identical. The Barlow loss is then based on the correlation matrix, rather than the cross-correlation matrix. The diagonal elements will always be one. Therefore, minimising the Barlow loss will require the off-diagonal elements to tend to zero, similar to orthogonal PLS latent variables.

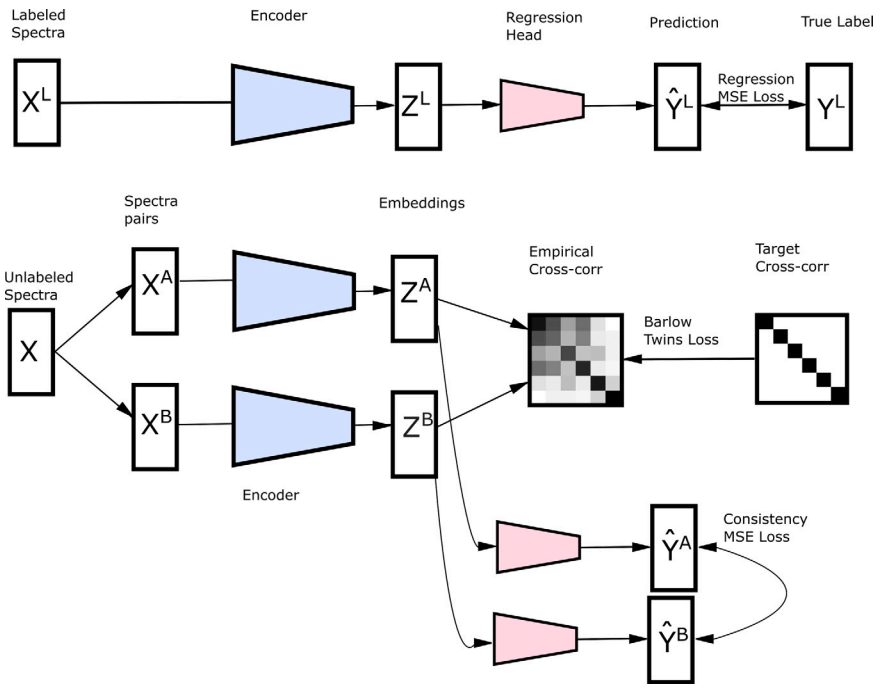

Fig. 2. Barlow Twins model for Spectral data. X represents unlabelled pairs of spectra from devices A and B, XL spectra from a separate labelled dataset, with labels YL (dry matter content or soluble solids content). The Barlow loss reduces redundancy and enforces invariance, the regression loss ensures that the network can predict Y, and the consistency loss ensures that predictions from spectra measured on the same fruit are similar. The regression head (pink) benefits indirectly from unlabelled data through the shared encoder (blue), which learns robust device invariant spectral representations from all paired measurements (including unlabelled data via the Barlow loss) while being constrained to produce accurate predictions on the labelled subset (via the MSE loss). During training, all three losses contribute to updating the encoder weights, with the weighted combination determining each loss's relative influence.

The other loss function, to be combined with the Barlow loss in a PLS-equivalent neural network approach, should maximise the covariance between the encoder outputs and the target variable. This approach is less straightforward and requires a custom loss function. For each latent variable, this is calculated as a partial covariance with the target variable, accounting for the previous latent variables.

4. Methods

4.1. Datasets

Our experiments employ a NIR dataset comprising kiwifruit across two sites in New Zealand (Te Puke and Kerikeri) and three seasons, with the last season being in 2019 season. Individual fruit were often measured by multiple devices. Only fruit with measurements from two or more devices were included in the study, resulting in a total of 4316 fruit and 10,869 scans over the 402 to 1137 nm range with a spectral resolution of 3 nm. All devices were of the same model, the F-750 Produce Quality Meter produced by Felix Instruments [18]. Reflectance values were measured with a xenon tungsten lamp. The measured range was trimmed to 222 wavelengths spanning 402–1065 nm and includes part of the visible (VIS, 402–780 nm) as well as the near-infrared (NIR, 780–1065 nm) regions. While wavelength selection methods are widely used in chemometrics to improve model performance [19], the wide range of wavelengths allows the proposed method to be assessed without specific feature engineering. Each fruit had its respective soluble solids content (SSC) and dry matter content (DMC) measured destructively. SSC was measured from a juice sample by a digital refractometer, expressed in °Brix. DMC represents the weight of a fruit tissue sample after drying as a percentage of its initial fresh weight. Summary statistics of these measures for the training and test sets are provided in Table 1. The first two years were used for the training tasks, with the remaining year used as a test set. No validation data was used for hyperparameter tuning. Five NIR spectrometers were used to collect the data, TP1, TP2, TP3 for the Te Puke site, and KK1, KK2 for the Kerikeri site.

Table 1
Summary statistics for dry matter (DMC) and soluble solids content (SSC) in training and testing sets.

Set	N	DMC				SSC			
		Mean	SD	Min	Max	Mean	SD	Min	Max
Training	2924	16.48	2.14	9.01	23.6	8.37	3.19	3.70	19.4
Testing	1392	17.48	2.48	10.30	24.2	7.69	3.12	3.75	19.7

Table 2
Number of fruit scanned per device in the training dataset. Diagonal elements indicate the total number of fruit measured by the respective device. Off-diagonal elements show the number of fruit measured by pairs of devices.

Device	KK1	KK2	TP1	TP2	TP3
KK1	**1397**	397	1089	90	90
KK2	397	**881**	90	574	573
TP1	1089	90	**1985**	986	986
TP2	90	574	986	**1616**	1615
TP3	90	573	986	1615	**1616**

The training set, comprising of 2924 fruit (7495 total scans), had some overlaps among the various devices. Table 2 shows the number of fruit scanned for each device (diagonal counts) as well as pairs of scans from different devices measured on the same fruit (offdiagonal). There was some mixing of the devices in the training set, but not in the 2019 data, where devices were restricted to a certain site (Table 3).

4.2. Model architecture

The model architecture used in these experiments was based on previous work on the same dataset. The encoder model consisted of an initial convolutional layer with bias set to zero and the weights

Table 3
Number of fruit scanned per device in the 2019 test dataset. The diagonal elements indicate the total number of fruit measured by the respective device. Off-diagonal elements show the number of fruit measured by pairs of devices.

Device	KK1	KK2	TP1	TP2	TP3
KK1	**797**	797	0	0	0
KK2	797	**798**	0	0	0
TP1	0	0	**592**	592	591
TP2	0	0	592	**594**	593
TP3	0	0	591	593	**593**

set to the Savitzky–Golay second derivative, with a window size of 13 and second-order polynomial. This layer was frozen so that it was equivalent to applying the respective Savitzky–Golay preprocessing. Following this, a convolutional layer with 50 filters of size 13 and a linear activation function was used as in [20]. The encoder was completed by flattening the convolutional layer and connecting a single linear dense layer of size 16. A regression head was then added to the model, which included a single linear neuron. The justification for the simple architecture was to enable a fair comparison between the Barlow method and PLSR, while avoiding confounding results with model complexity.

4.2.1. PLSR model

The PLSR regression was fit using scikit learn [21] and used Savitzky–Golay second derivative, with a window size of 13, and a second-order polynomial. All models used 16 latent variables to provide a direct comparison with the neural network encoder's architecture. This choice was validated using 10-fold cross validation (CV) PLSR models across different training set sizes. For SSC prediction, 16 components minimised the CV RMSE for moderate training sizes (N = 250 fruit, 648 scans; N = 500 fruit, 1297 scans), with 17 optimal at N = 1000 fruit (2122 scans). For DMC prediction, cross validation selected 20 and 21 latent variables at N = 100 and N = 250, respectively, but evaluating on the test set showed optimal numbers of latent variables to be 12 and 13, with 16 latent variables giving similar performance. The small training datasets (50 fruit, 150 scans) had a slightly lower optimal (13 latent variables) for the SSC and DMC prediction. The consistent use of 16 latent variables across experiments ensures that architectural changes to the models did not affect the Barlow Twins loss's performance.

4.3. Training procedure

All analyses were conducted in Google Colab notebooks with a T4 GPU and high RAM backend. Software used included Python 3.11.13, Tensorflow 2.18.0, Tensorflow Probability 0.25.0, and scikit-learn 1.6.1. Models were trained for 1000 epochs using the Lamb optimiser [22] with a learning rate of 0.01.

A custom loss function was used for training with a weighted combination of 10 x Barlow loss + 0.5 consistency MSE loss + 0.5 prediction MSE loss. The λ parameter in the Barlow loss was set to 1/15 (the number of latent variables minus one), so that the invariance and redundancy terms in Eq. (1) are equally weighted. For n latent variables, the $n \times n$ cross-correlation matrix has n diagonal elements and $n(n-1)$ off-diagonal elements. For the experiments, the Barlow loss was computed using only the labelled subset of data (Barlow Labelled), or using all available paired device measurements (Barlow Full), regardless of whether the fruit had DMC or SSC labels.

4.4. Experiments

4.4.1. Semi-supervised learning

To investigate the effectiveness of semi-supervised learning we used different labelled training set sizes (50, 100, 250, 500, 1000, 2924 fruit) for separate models predicting DMC and SSC. These fruit were selected by taking the n most recent fruit and their respective scans in the training set. This approach was more realistic than sampling a random subset of scans from the full training set. The full 2924 fruit training set (7495 scans) was used as unlabelled data to train the model encoder using the Barlow and consistency loss functions. This semi-supervised approach (Barlow Full) was compared to a the same model with the Barlow loss calculated only on the subset of labelled fruit (Barlow Labelled, using the paired device measurements available for those fruit, rather than on the full unlabelled dataset. Both methods combined the Barlow loss with the regression MSE loss, but differed in the size of the dataset used for computing the Barlow loss. Similarly, autoencoder models based on methods from [13] were trained on the same datasets (AE Full and AE Labelled). Additionally, models trained with the MSE loss only and PLS regression with Savitzky–Golay preprocessing were included as baselines.

Six modelling approaches were compared based on their loss functions and use of unlabelled data:

- **Barlow Full:** Barlow Twins loss computed on full training set measurements (unlabelled) + MSE loss on labelled subset
- **Barlow Labelled:** Barlow Twins loss computed only on labelled paired measurements + MSE loss on labelled subset
- **AE Full:** Autoencoder reconstruction loss computed on full training set measurements (unlabelled) + MSE loss on labelled subset
- **AE Labelled:** Autoencoder loss computed only on labelled paired measurements + MSE loss on labelled subset
- **MSE:** Standard supervised learning with MSE loss on labelled subset only
- **PLSR:** 16-component PLS regression with Savitzky–Golay preprocessing

4.4.2. Calibration transfer

Calibration transfer experiments involved excluding one device at a time from the training data. In each iteration, the excluded device served as the target device for testing, with the remaining four serving as source devices for training. The scans from the left-out device were completely excluded from calculating the Barlow loss to prevent data leakage.

Models were fit using different loss functions across increasing labelled training set sizes (N = 50, 100, 250, 500, 1000, 2924 fruit):

- **Barlow Full:** Barlow Twins loss computed on full training set measurements from source devices (unlabelled) + MSE loss on labelled subset
- **Barlow Labelled:** Barlow Twins loss computed only on labelled paired measurements + MSE loss on labelled subset
- **MSE:** Standard supervised learning with MSE loss on labelled subset only
- **PLSR:** 16-component PLS regression with Savitzky–Golay preprocessing

Model performance was assessed on the 2019 test data for the respective left out device and aggregated across all devices.

4.4.3. Augmentation

The original Barlow Twins paper used image augmentation to train the network. This is a viable alternative when collected data is only available for a single device. The augmentation method implemented here involves adding random samples from a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the empirical covariance of differences among devices on kiwifruit measured prior to the current dataset on a wider range of devices. The dimension of $\boldsymbol{\Sigma}$ is equal to the number of wavelengths included in the model training. The augmented spectrum is then calculated as:

$$\mathbf{x}' = \mathbf{x} + \boldsymbol{\varepsilon}, \quad \text{where } \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{3}$$

Table 4
Comparison of different loss functions tested on the 2019 data across different Training Set Sizes. N represents the number of labelled fruit samples in the training set. Full Data uses all training spectra (labelled and unlabelled) for calculating the Barlow and Autoencoder losses, while Labelled uses only the labelled spectra for these losses. All RMSE values are calculated on the same 2019 labelled test set.

Training	Measure	2019 test RMSE					
		Full data		Labelled data only			
N		AE	Barlow	AE	Barlow	MSE	PLS
50	DMC	4.23	**1.55**	3.45	1.69	1.60	1.87
100	DMC	3.19	**1.33**	3.72	2.17	2.45	1.54
250	DMC	2.18	**1.17**	2.05	1.21	1.27	1.27
500	DMC	2.40	1.20	2.14	1.24	**1.17**	1.38
1000	DMC	3.97	1.19	1.84	**1.15**	**1.15**	1.42
2924	DMC	1.31	1.17	1.58	1.14	**1.11**	1.42
50	SSC	3.19	2.07	2.77	1.97	4.61	**1.74**
100	SSC	2.91	2.26	2.28	**2.22**	2.91	2.25
250	SSC	2.29	1.98	2.34	1.92	2.03	**1.70**
500	SSC	2.73	1.96	2.32	1.84	2.22	**1.42**
1000	SSC	2.15	1.65	2.42	1.63	1.70	**1.39**
2924	SSC	1.43	1.49	1.64	**1.43**	1.46	1.62

See [17] for more details. If the Barlow loss is used for training on unlabelled scans from a single device, it is effectively only enforcing the latent variables to be orthogonal and has no invariance to device variation. For these experiments, the model was fitted on data from a single device, while augmenting the unlabelled spectra to train the encoder using the Barlow loss function and MSE loss to train the regression head on the labelled data. For comparison, equivalent models trained using only the labelled data were also created. Model performance was assessed on the 2019 test data for the respective device and all other devices separately.

5. Results

5.1. Semi supervised learning

Table 4 summarises the performance on the 2019 data of semi-supervised (SSL) and supervised learning (SL) approaches using different loss functions (Barlow, AE, MSE) compared to a PLSR baseline.

Generally, all models improved as the number of fruit included in the labelled training set increased. The main exception was for the autoencoder loss, which performed the worst until the full training set was used. However, at $N = 2924$ fruit, the SL and SSL methods are equivalent and there is considerable differences observed (RMSE 1.31 and 1.43 compared to 1.58 and 1.64 respectively), indicating it may be less stable than the Barlow loss.

Among the neural network approaches, the Barlow loss consistently achieved the best performance, especially with the smaller training sets. For large training sizes, there appeared to be no negative impact to using the Barlow loss, with the test RMSE being practically identical to that of training the same model using only the MSE loss. The SSL method with the Barlow loss reduced the RMSE at the smaller training sets for the DMC, with SSL and SL results converging as the sample size increases. This is expected as the data used to calculate the Barlow loss in the SL model is a subset of the full data used in the SSL.

These RMSE reductions were consistent across the range of observed DMC and SSC values.

Compared to the PLS baseline, the Barlow and MSE models outperformed PLSR for DMC. Interestingly, when training on 100 labelled fruit instead of 50, the RMSE increased for the SL neural networks, but not PLSR or the SSL neural networks. For SSC, PLSR had the lowest RMSE until the full training set was used.

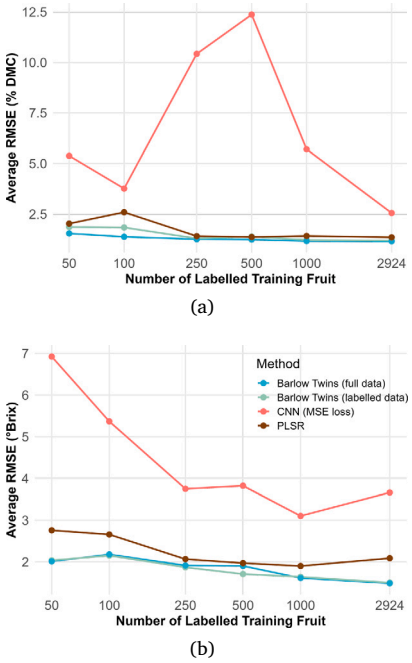

Fig. 3. Calibration transfer results for DMC (a) and SSC (b) prediction using leave-one-device-out validation. Models were trained on four source devices and tested on the remaining target device using different loss functions and increasing numbers of labelled fruit in the training set. RMSE is averaged across all five target devices. Barlow Twins (full data) uses all training set data for the Barlow loss, while Barlow Twins (labelled data) uses only the labelled data.

5.2. Calibration transfer

The results of leaving a single device out of the training set and testing on that respective device are summarised in Figs. 3(a) and 3(b). As expected, the RMSE generally decreased across the models as the number of fruit included in the training set increased.

For both DMC and SSC prediction the Barlow loss models consistently outperformed the PLSR baseline and prevented the extremely poor performance of the equivalent CNN with the MSE loss only. This CNN showed instability at the low to moderate training sizes with RMSE exceeding 12% for DMC at N = 500 fruit and 7 °Brix for SSC at N = 50 fruit despite low training losses, indicating overfitting to source devices. In contrast, both Barlow models were stable across all labelled training sizes.

Including the full data for the Barlow Twins loss showed some improvement over using the labelled data only (Fig. 3(a)) with lower RMSE for the smaller training set sizes (50 and 100 fruit). However, this was not observed for the SSC prediction (Fig. 3(b), where the two Barlow Twins models showed practically identical performance across all the training sizes. The Barlow Twins loss provides regularisation by constraining the structure on the encoder embeddings to have low redundancy in the latent variables and high device invariance, similar to PLSR latent variables. Neither of these methods had the extreme RMSE observed with the less constrained MSE training.

5.2.1. Augmentation

Results of training on augmented spectra from a single device and predicting on 2019 data from the same device are shown in Figs. 4(a) and 4(c). Results for predicting on the remaining four devices not used in training are shown in Figs. 4(b) and 4(d). Contrary to previous results, there appears to be little benefit in using the full unlabelled training set for a given device when minimising the Barlow loss over

using the labelled data only. The RMSE values for the two Barlow Twins methods are consistently similar across the two measures and training sizes. The only notable exception is SSC at N = 100 fruit when predicting on the same device, which appears to be driven by a single device at that training size(data not shown). The similarity in Barlow Twins results may be explained by the augmentation used, adding an independent random sample from a multivariate Gaussian distribution. Since this is independent of the observed spectra being added to, there is less to gain in increasing the unlabelled data than if the augmentation depended on the observed spectra. More complex augmentation techniques that are related to the observed spectra may benefit from semi-supervised techniques.

The Barlow loss function provided, whether used with the labelled data only or the additional unlabelled data, saw lower RMSE than training with the MSE loss function at the smaller training sizes. This was more pronounced when testing on devices left out of training (Figs. 4(b) and 4(d)).

6. Discussion

6.1. Advantages over traditional approaches

Barlow Twins improves upon PLSR by explicitly learning device-invariant representations. PLSR learns latent variables that maximise covariance with the target while enforcing orthogonality. However, it lacks a specific mechanism to ensure that these latent variables are invariant to device variability. The Barlow loss simultaneously reduces redundancy (like PLSR) while also enforcing that paired measurements of the same fruit produce similar embeddings, regularising the network to prevent overfitting to device specific features. This overfitting is observed in the calibration transfer results in Fig. 3, where the CNN trained with the MSE loss overfits to the training devices, producing the highest RMSE for both SSC and DMC. In contrast, the models that included the Barlow loss were stable because of the invariance constraint, which penalises device-specific encodings. As the labelled data increase, PLSR's regularisation via dimensionality reduction and covariance maximisation becomes sufficient, and the benefits of semi-supervised learning diminish (Table 4).

6.2. Limitations and challenges

The current research focused on the Barlow Twins method in a semi-supervised method, without optimising the architecture of the encoder or regression head. The results may vary for a more flexible architecture. Limited experiments with deeper encoders encountered training issues with exploding gradients, although this was overcome by using the SELU [23] activation function and LeCun norm initialisation for the encoder.

A potential issue is that fixing the number of latent variables to 16 may disadvantage PLSR models. We investigated this by repeating the experiments in Table 4 using 10-fold CV to select the optimal number of PLSR latent variables. The CV-optimised models showed performance similar to, or slightly worse than, that of the models with 16 components. Some improvement was observed at the largest training size, where CV selected more latent variables, but this did not change the overall conclusions.

Another limitation is the augmentation used for NIR spectra; we employ the method from [17], but alternative augmentation methods could also be considered. Temperature variations, for example, have been shown to be associated with wavelength shift in raw milk [24]. Augmenting the data by wavelength shifting for use with this model may improve model sensitivity to these effects.

DMC and SSC may vary throughout the fruit depending on the tissue location sampled. The NIR measurements were taken at consistent locations to minimise this variation. The success of the Barlow

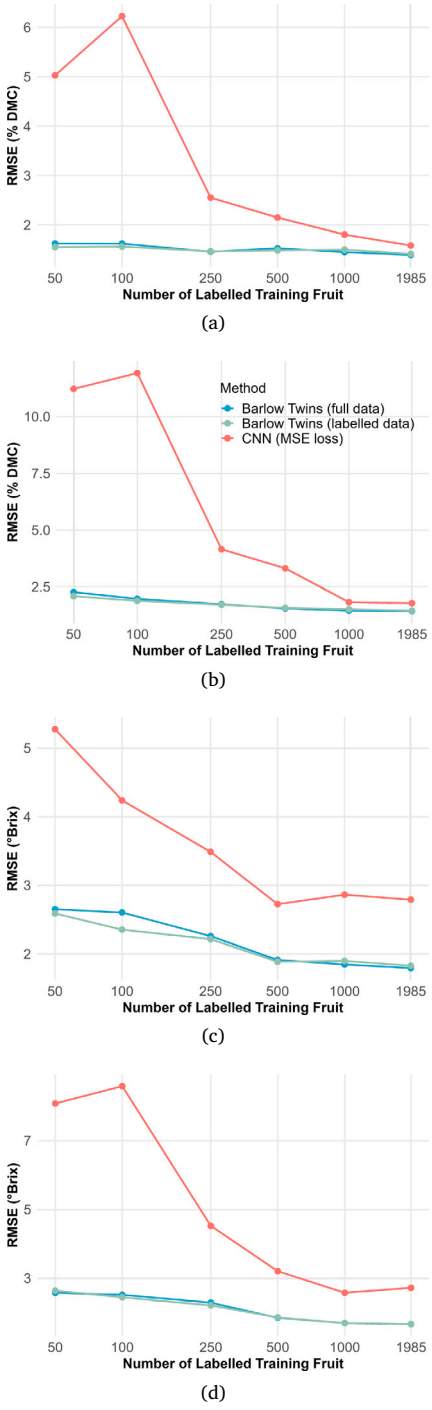

Fig. 4. Models trained with different loss functions on augmented spectra from a single device. Models were evaluated on 2019 data for the same device (DMC in (a), SSC in (c)) and from the remaining four devices (DMC in (b), SSC in (d)). RMSE is averaged across the five training devices with varying numbers of labelled fruit. Barlow Twins (full data) uses all training set data for the Barlow loss, while Barlow Twins (labelled data) uses only the labelled data. The rightmost point (1985) represents the maximum available training data, which varied by device (881 to 1985 fruit).

Twins method suggests that the paired measurements were based on sufficiently similar tissue.

While we investigated the Barlow Twins approach, other semi-supervised learning approaches could also perform well. The comparison to the autoencoder method used in [13] was limited due to the lack of details of their model architecture. The autoencoder employed here was based on a separate paper looking at NIR spectra from the same brand of devices as used here [10], but there may be alternative autoencoder architectures better suited to kiwifruit. Other self-supervised methods could also improve performance. For example, variance–invariance–covariance regularisation (VicReg) [25] is a similar method to Barlow Twins. While preliminary experiments (not shown) with this did not improve over the Barlow Twins method, a more thorough investigation may yield improved results.

6.3. Practical implications

Large quantities of unlabelled data may not always be available, but the method is reasonably straightforward to implement and showed benefits even with smaller datasets. The computational overhead is modest and mainly due to the increased number of calculations involved in comparing all pair-wise device combinations of the same fruit for the Barlow Twins loss. For optimal results, unlabelled data should be collected in a systematic way to improve robustness to various effects. This could be multiple scans on the same fruit at different temperatures, multiple devices scanning the same fruit (as observed here), or multiple measurements at different locations on the fruit, as is seen in commercial fruit graders.

7. Conclusions

The results show the potential benefit of self-supervised learning for NIR spectroscopy, through adapting the Barlow Twins method for regression tasks. The greatest benefit over traditional PLS regression in terms of RMSE is consistently observed when the labelled data is limited. The robustness of the method to device variation could be particularly useful in practice. Labelling enough spectra to train a new device using standard supervised methods, such as PLSR, can be time-consuming and expensive. For example, each kiwifruit DMC measurement requires two equatorial slices to be dried for 24 h at 65 °C [26]. Conversely, spectral measurements of the same kiwifruit are simple and quick, taking about 4–6 s per scan for the Felix f-750 handheld devices used to collect the data in this study [27].

Future work could explore alternative self-supervised learning methods, find optimal model architectures for this application, and appropriate augmentation methods for the model to learn invariance. Additionally, measuring the same fruit from the same device multiple times would allow for an investigation into whether methods such as the one presented in this paper are able to reduce measurement error.

CRediT authorship contribution statement

Mark Wohlers: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andrew McGlone:** Writing – review & editing, Supervision. **Eibe Frank:** Writing – review & editing, Supervision. **Geoffrey Holmes:** Writing – review & editing, Supervision.

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.

Acknowledgements

The research received financial support by the New Zealand Ministry of Business Innovation & Employment (MBIE) as part of the Endeavour funded project *Perfecting storage life prediction for delivery of high quality fruit*.

Data availability

All data used in this paper are available through Wohlers et al. 2023 [17]. Code is available at https://github.com/mwohlers/BarlowTwins_NIR.

References

[1] C. Cui, T. Fearn, Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration, Chemometr. Intell. Lab. Syst. 182 (2018) 9–20, http://dx.doi.org/10.1016/j.chemolab.2018.07.008.

[2] J. Walsh, A. Neupane, A. Koirala, M. Li, N. Anderson, Review: The evolution of chemometrics coupled with near infrared spectroscopy for fruit quality evaluation II. The rise of convolutional neural networks, J. Compunetics Infrared Spectroscop 31 (3) (2023) 109–125, http://dx.doi.org/10.1177/09670335231173140.

[3] E.J. Bjerrum, M. Glahder, T. Skov, Data augmentation of spectral data for convolutional neural network (CNN) based deep chemometrics, 2017, http://arxiv.org/abs/1710.01927.

[4] P. Mishra, D. Passos, F. Marini, J. Xu, J.M. Amigo, A.A. Gowen, J.J. Jansen, A. Biancolillo, J.M. Roger, D.N. Rutledge, A. Nordon, Deep learning for near-infrared spectral data modelling: Hypes and benefits, TRAC Trends Anal. Chem. 157 (2022) 116804, http://dx.doi.org/10.1016/J.TRAC.2022.116804.

[5] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Automatic differentiation in pytorch, 2017, https://openreview.net/forum?id=BJJsrmfCZ.

[6] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, R. Monga D. Mané, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, O. Vinyals F. Viégas, P. Warden, M. Wattenfberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, Version 1 (15) (2019) https://github.com/tensorflow/tensorflow/releases/tag/v1.15.0.

[7] D. Makalesi, Çataltaş, K. Tütüncü, Y. Kızılötesi, S. Kullanlan, V. Analizi, T. Bir, D. Öz, A review of data analysis techniques used in near-infrared spectroscopy, Eur. J. Sci. Technol. 25 (2021) 475–484, http://dx.doi.org/10.31590/ejosat.882749, http://dergipark.gov.tr/ejosat475.

[8] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometr. Intell. Lab. Syst. 58 (2) (2001) 109–130, http://dx.doi.org/10.1016/S0169-7439(01)00155-1, https://www.sciencedirect.com/science/article/pii/S0169743901001551.

[9] J. Gui, T. Chen, J. Zhang, Q. Cao, Z. Sun, H. Luo, D. Tao, A survey on self-supervised learning: Algorithms applications, and future trends, IEEE Trans. Pattern Anal. Mach. Intell. (2024) http://dx.doi.org/10.1109/TPAMI.2024.3415112.

[10] L. Zhang, J. Liu, Y. Wei, D. An, X. Ning, Self-supervised learning-based multi-source spectral fusion for fruit quality evaluation: A case study in mango fruit ripeness prediction, Inf. Fusion 117 (2025) 102814, http://dx.doi.org/10.1016/J.INFFUS.2024.102814, https://www.sciencedirect.com/science/article/pii/S156625352400592X?ssrnid=4903512&dgcid=SSRN_redirect_SD.

[11] J. Zbontar, L. Jing, I. Misra, Y. LeCun, S. Deny, Barlow twins: Self-supervised learning via redundancy reduction, Proc. Mach. Learn. Res. 139 (2021) 12310–12320, https://arxiv.org/pdf/2103.03230.

[12] J.E. van Engelen, H.H. Hoos, A survey on semi-supervised learning, Mach. Learn. 109 (2) (2020) 373–440, http://dx.doi.org/10.1007/S10994-019-05855-6/FIGURES/5, https://link.springer.com/article/10.1007/s10994-019-05855-6.

[13] M. Said, A. Wahba, D. Khalil, Semi-supervised deep learning framework for milk analysis using NIR spectrometers, Chemometr. Intell. Lab. Syst. 228 (2022) 104619, http://dx.doi.org/10.1016/J.CHEMOLAB.2022.104619, https://www.sciencedirect.com/science/article/abs/pii/S0169743922001307.

[14] P. Mishra, E. Woltering, Semi-supervised robust models for predicting dry matter in mango fruit with near-infrared spectroscopy, Postharvest Biol. Technol. 200 (2023) 112335, http://dx.doi.org/10.1016/J.POSTHARVBIO.2023.112335, https://www.sciencedirect.com/science/article/pii/S0925521423000960.

[15] M. Dhaini, M. Berar, P. Honeine, A. Van Exem, Contrastive learning for regression on hyperspectral data, 2024, https://arxiv.org/pdf/2403.17014v1.

[16] B. Hapke, Bidirectional reflectance spectroscopy: 1 theory, J. Geophys. Res.: Solid Earth 86 (B4) (1981) 3039–3054, http://dx.doi.org/10.1029/JB086IB04P03039, /doi/pdf/10.1029/JB086iB04p03039.

[17] M. Wohlers, A. McGlone, E. Frank, G. Holmes, Augmenting NIR Spectra in deep regression to improve calibration, Chemometr. Intell. Lab. Syst. 240 (2023) 104924, http://dx.doi.org/10.1016/J.CHEMOLAB.2023.104924.

[18] Felix Instruments, F-750 produce quality meter, \ https://felixinstruments.com/food-science-instruments/nir-spectroscopy/f-750-produce-quality-meter/, 2019, https://felixinstruments.com/food-science-instruments/nir-spectroscopy/f-750-produce-quality-meter.

[19] G.H. Yamashita, M.J. Anzanello, F. Soares, M.K. Rocha, F.S. Fogliatto, Selecting relevant wavelength intervals for PLS calibration based on absorbance interquartile ranges, Chemometr. Intell. Lab. Syst. 231 (2022) 104689, http://dx.doi.org/10.1016/J.CHEMOLAB.2022.104689.

[20] M. Wohlers, A. McGlone, E. Frank, G. Holmes, Assessing machine learning models for near-infrared regression by measuring stability towards diffeomorphisms, Chemometr. Intell. Lab. Syst. 264 (2025) 105449, http://dx.doi.org/10.1016/J.CHEMOLAB.2025.105449, https://www.sciencedirect.com/science/article/pii/S0169743925001340.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830, http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.

[22] Y. You, J. Li, S. Reddi, J. Hseu, S. Kumar, S. Bhojanapalli, X. Song, J. Demmel, K. Keutzer, C.J. Hsieh, Large batch optimization for deep learning: training BERT in 76 minutes, in: 8th International Conference on Learning Representations, ICLR 2020, 2019, https://arxiv.org/pdf/1904.00962.

[23] G. Klambauer, T. Unterthiner, A. Mayr, S. Hochreiter, Self-normalizing neural networks, Adv. Neural Inf. Process. Syst. 2017-December (2017) 972–981, https://arxiv.org/pdf/1706.02515.

[24] J.A. Diaz-Olivares, S. Grauwels, X. Fu, I. Adriaens, W. Saeys, R. Bendoula, J.M. Roger, B. Aernouts, Temperature correction of near-infrared spectra of raw milk, Chemometr. Intell. Lab. Syst. 255 (2024) 105251, http://dx.doi.org/10.1016/J.CHEMOLAB.2024.105251, https://www.sciencedirect.com/science/article/pii/S0169743924001916#fig1.

[25] A. Bardes, J. Ponce, Y. LeCun, VICReg: variance-invariance-covariance regularization for self-supervised learning, in: ICLR 2022-10th International Conference on Learning Representations, 2021, https://arxiv.org/pdf/2105.04906.

[26] V.A. McGlone, S. Kawano, Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy, Postharvest Biol. Technol. 13 (2) (1998) 131–141, http://dx.doi.org/10.1016/S0925-5214(98)00007-6, https://www.sciencedirect.com/science/article/pii/S0925521498000076#FIG1.

[27] FreshView. https://www.freshview.com.au/page/felix_750.