Analytica Chimica Acta 1191 (2022) 339308

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Deep generative neural networks for spectral image processing

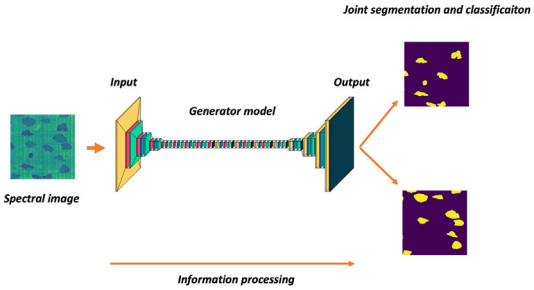
Puneet Mishra

Wageningen University & Research, Food and Biobased Research, Wageningen, the Netherlands

HIGHLIGHTS

- Image translation was proposed for spectral image processing.
- The generative adversarial network was used for image translation.
- The image translation uses both spatial and spectral information.
- The applications such as segmentation, regression and classification are demonstrated.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 24 September 2021
Received in revised form
15 November 2021
Accepted 18 November 2021
Available online 21 November 2021

Keywords:
Neural networks
Spatial-spectral
Generative models
Spectroscopy

ABSTRACT

An artificial intelligence approach based on deep generative neural networks for spectral imaging processing was proposed. The key idea was to treat different spectral image processing operations such as segmentation, regression, and classification as image-to-image translation tasks. For the image-to-image translation, the conditional generative adversarial networks were used. As a baseline comparison, the traditional chemometric approach based on pixels wise modelling was demonstrated. The analysis was presented with two real data sets related to fruit property prediction and kernel and shell classification of walnuts. The presented artificial intelligence approach for spectral image processing can provide benefits for any field of science where spectral imaging and processing is widely performed.

1. Introduction

Spectral imaging is a bi-modality analytical technique where, on one mode, the imaging captures the spatial information about the samples and, on another mode, the spectroscopy captures spectral characteristics of samples under study [1,2]. Both modes combined capture the spatially resolved spectral properties of samples [3]. Spectral imaging is one of the most widely used analytical techniques for non-invasive, non-contact analysis as well as it requires minimal sample preparation and provides the analysis results in real-time once pre-calibrated [1,3–5]. Applications of spectral imaging can be found ranging from agriculture [6,7] and foods [1] to high-end pharmaceuticals [8]. Based on the experimental needs, spectral imaging can be explored for a range of the electromagnetic spectrum such as ultraviolet [9], visible and near-infrared [10,11], mid-infrared [12], terahertz [13] etc. In the literature, several applications of Raman spectral imaging [14] can also be identified.

Although the current state of the art presents diverse sensing options to spectral imaging, spectral imaging alone is of no use unless it is combined with advanced data processing and modelling approaches to extract meaning out of the rich spatial-spectral data captured by the spectral camera [3,4]. The data modelling is required to teach the cameras to be able to recognise key analytes,

E-mail address: puneet.mishra@wur.nl.

https://doi.org/10.1016/j.aca.2021.339308

as these cameras have no inbuilt intelligence. In the domain of chemometrics, spectral image processing is widely performed using unsupervised and supervised data modelling approaches [3,4]. The popular chemometric approach for spectral image processing involves extracting key subset spectra from the spectral images using a region of interest (ROI) selection [3]. Later, the traditional spectral data modelling is performed in the subset of spectra using approaches such as principal component analysis, clustering and partial least-squares based regression and classification, depending on the need for unsupervised data exploration or supervised predictive modelling [4]. As a third step, the developed models are used to predict the modelled property for each pixel of the image to achieve the commonly known prediction maps [3,4]. Further steps can involve post-processing of the prediction maps for specific tasks such as object location, bounding box operation, counting, and many more. Applications of such pixel-wise modelling approaches can be found widely in the scientific literature [15–21].

Although the pixel-wise modelling approach involving chemometric methods is widely used in the spectral image processing domain, there are also drawbacks involved with such an approach. The main drawback is that the pixel-wise modelling approach does not treat the spectral images as images but treat them like a large point spectroscopy database [22–24]. This is because the modelling on a subset of pre-selected spectra leads to loss of the spatial information present in the image [6,7]. Usually, the spatial information in the scene carries contextual information which in the spectral image processing domain has been proven to be of high value to improve the model performance [22]. This is also one of the main reasons for the increasing interest of the chemometric community in the usage of spatial information alongside spectral information during spectral image processing [25–28]. Nonetheless, the current approaches used by the chemometric community to incorporate spatial information into the models still involve a low-level fusion of spatial and spectral information before pixel-wise chemometric modelling [27,28]. In such a low-level fusion, usually spatial filters are used to extract the textural features from the images. These extracted texture features are then stacked behind the spectral planes and the image is processed in the traditional pixel-wise approach [27,28]. Several studies have shown that combining textural features with spectral information improved the model's performance [27–30]. Such an approach allows incorporating the textural features from the imaged scene into the models, however, it still does not consider the contextual information present in the imaged scene. This is because, after a low-level fusion of the spectral and textural features, the spectral images are still processed in a pixel-wise approach, hence the spatial contextual information is underutilized.

Going outside the chemometric domain and to the computer vision domain where massive development has taken place in terms of deep learning (DL) approaches to image processing, it can be noted that contextual information is largely used during image processing and model development. More than the spectral dimensions, which are the red, green, and blue colour bands in a colour image, the most useful information is the contextual information present in the imaged scene (shapes, textures, etc.) [31]. In the DL domain, to take advantage of the contextual information, convolutional neural networks (CNNs) are usually used, wherein the convolutional layers allow modelling and extracting of the rich spatial contextual information, while the rest of the neural network (usually dense layers) allows mapping the extracted features with the target property [32]. The CNNs are also increasingly gaining traction for spectral image processing, more generally for supervised classification purposes, where spectral images are assigned to a particular class [33,34]. Recently, several recent works have also shown the potential of fully convolutional neural networks such as U-net [6,7] and generative adversarial networks (GANs) [22] for semantic segmentation of spectral images. However, until now there is no work that has explored the CNNs for dealing with the semantic image regression and classification of the spectral images. Regression and classification are two main tasks performed in spectral image processing, where for multiple objects present in the imaged scene either the reference property needs to be estimated (regression) or, the objects need to be assigned to the right classes (classification) [3]. This study, for the first time, puts forward the concept of image translation for semantic regression and classification of the objects present in spectral images. In easy terms, the image translation can be understood as tasks such as colouring a black and white image and transforming satellite images to road maps [35]. A powerful approach to image translation in the DL domain is the GAN, which involves an adversarial training of a combination of generator and discriminator neural network models [36]. The generator model task is to learn to perform image translation (for example, developing prediction maps for spectral images) and the discriminator model task is to detect if the generator model can translate the images efficiently. At a certain moment during the adversarial training, the generator and discriminator model achieve equilibrium where the generator model could translate images of plausible quality such that it becomes difficult for the discriminator model to detect if the images generated are the ground truth or the synthesized by the generator. Recent application of GAN models for semantic segmentation of spectral images [22] has shown promising potential for its further exploration toward semantic regression and classification of objects present in the spectral images.

The study aimed to present a new artificial intelligence approach to spectral image processing. The key idea was to demonstrate the image-to-image translation for different spectral image processing operations such as segmentation, regression and classification for objects present in the spectral images. For the image-to-image translation, conditional GANs (cGANs) were used. As a baseline comparison, the traditional chemometric approach based on pixels-wise modelling using the partial least-squares (PLS) analysis was demonstrated. All the data used in this study come from real data sets measured with an Fx10 (Specim, Finland) spectral camera. Furthermore, the effects of combining chemometric approaches such as pre-processing and dimensionality reduction with cGANs modelling were also explored.

2. Materials and methods

2.1. Samples and spectral imaging

In this study, to demonstrate the artificial intelligence way to spectral image processing two real data sets were gathered. The first data set, which was used to demonstrate the semantic segmentation and regression task to predict soluble solids content (SSC), was measured on 100 individual black grapes acquired from a local supermarket (AlbertHeijn, Ede, The Netherlands). For the classification case, walnut shells and kernels were used, where the task was to classify the walnut shells and kernels into separate classes. The walnuts (with shell) were also purchased from a local supermarket (Jumbo, Wageningen, The Netherlands). Prior to the experiment, the walnuts were manually opened using a nutcracker.

All images were captured with the All-in-One spectral imaging (ASI) setup as shown in Fig. 1. The ASI setup has an integrated visible and near-infrared spectral camera (Fx10, Specim, Oulu, Finland). The illumination is provided with two sets of halogen lights (supplied by Specim, Oulu, Finland) mounted next to the camera. The ASI setup is fully automated for image acquisition, and the acquisition controls such as the speed of the translation stage,

Fig. 1. The all-in-one spectral imaging setup used for data acquisition. This all-in-one [37] setup provides a fully automated approach to spectral imaging and the final output of the setup is the raw reflectance data.

exposure time, number of frames, etc. are synchronised with the camera settings. The ASI setup has an inbuilt white reference (Teflon) which is scanned prior to any image acquisition. All images are automatically corrected for the white and dark reference and the final output of the ASI setup is the reflectance spectral data which can be directly used for any data modelling task.

A key point to note is that for the regression analysis there was a need for a reference property to be regressed, hence, in this study the soluble solids content (SSC) of the grape juice was used. SSC of extracted fruit juice was determined using a handheld refractometer (HI 96801, Hanna Instruments Inc, Woonsocket, RI, USA).

2.2. Data sets generated

The image acquisition resulted in a total of four spectral images of size 900 × 1024 × 224 where the first two dimensions were the spatial dimension, and the last dimension was the spectral sampling in the wavelengths range of 398−1002 nm. Out of the four images, two images belonged to the grapes data set where each image consists of measurements on 50 grape samples. The other two images were from the walnut shells and kernels, which were randomly distributed on the imaging platform in the ASI setup (Fig. 1).

2.3. Data pre-processing

The spectra in the spectral range below 471 nm were noisy; hence, the spectral range was reduced to 471−1002 nm, consisting of 195 spectral bands. A point to note is that the edges in the images containing no information about the samples were cropped to reduce the spatial dimensions. Such a reduction was performed because it reduces the size of the data set which can be handled more easily during deep model training. Later, all spectral images were smoothened in the spectral domain using the Savtizky-Golay [38] filter (window width = 13 and polynomial order = 2). The smoothing in the spectral domain was performed to reduce any subtle variation in the spectra. In this study, the effect of spectral normalization was also explored on the DL models, hence, the spectra were also pre-processed with the standard normal variates (SNV) [39]. Notice that the DL was performed separately on both the normalised and data without any normalization. In this study, the effect of data compression prior to deep learning was also explored, and for that purpose, the partial least-squares (PLS) based latent space modelling was used. The PLS based latent space transformation was performed separately for raw and SNV normalised data. For all the cases, the spectral data were transformed to score maps by multiplying the spectral image with the loading vectors extracted from the PLS analysis. In the case of the grape data

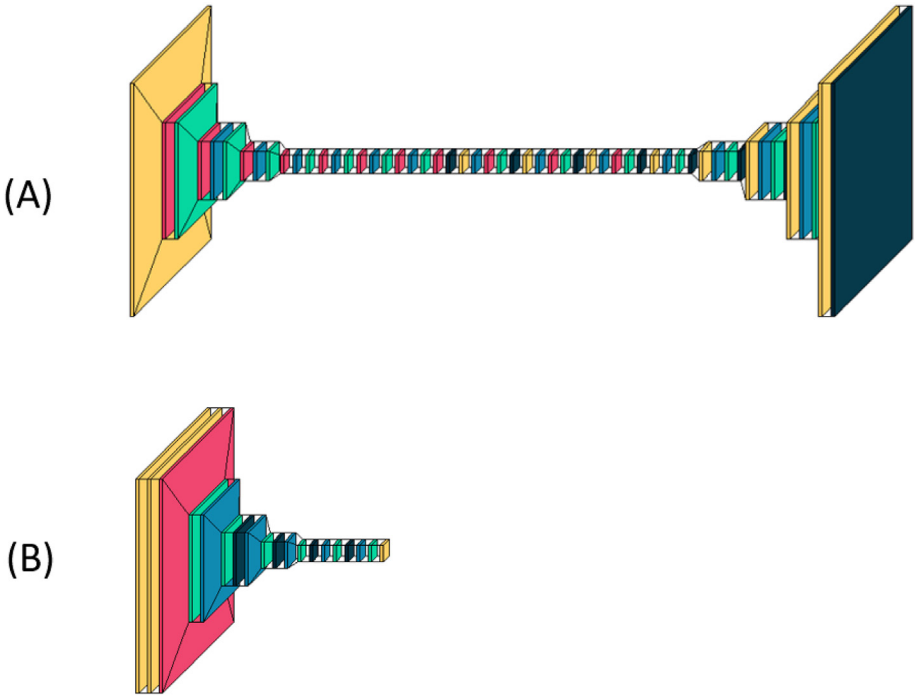


Fig. 2. A schematic of the generator (A) and discriminator model (B). The "U-net" based generator model has an encoder part: C64−C128−C256−C512−C512−C512−C512−C512 and a decoder part: CD512-CD512-CD512-C512-C256-C128-C64, where C represents a convolution, CD means transpose convolution and the number mean convolutional filters. The discriminator model architecture was: C64−C128−C256−C512, where C stands for convolution and the number stands for convolutional filters.

set, the optimal number of latent variables (LVs) was identified by regressing the mean spectra of the grapes with the SSC values and using 10-fold Venetian blind cross-validation. For the walnut shells and kernels classification, the optimal number of LVs were identified by regressing a small set of spectra (~144 for walnut shells and 144 spectra for nuts) extracted from the images using a manual selection which were later regressed with a one-hot-encoded vector of [111 … 000], where 1 signifies the walnut shells and 0 signifies the kernels. The optimal LVs for the walnut data set were also found using 10-fold Venetian blind cross-validation. The optimal LVs were identified as the elbow point in the cross-validation plots. Finally, the DL analysis was performed independently on the four sets of data i.e., raw reflectance data, SNV normalised data, PLS transformed raw reflectance data, and PLS transformed SNV data. A key point to note is that the reflectance data were on the scale of [0−1], while after the SNV transform and after the PLS projection the data scale was transformed. The neural networks models require data to be in [0−1] to work efficiently. Hence, the SNV and PLS transformed data were rescaled between [0−1] by first adding the lowest negative value for that spectral plane and then dividing by the max values for each spectral plane independently. The minimum and maximum values were stored as during the application of the model on the new data set, these values were used to scale the images.

2.4. Ground truth maps

In a typical supervised learning task such as segmentation, regression and classification, there is a need to account for the reference values against which the model can be trained, validated and tested in its performance. For DL modelling, there is also a need to account for ground truth reference values against which the DL model can be trained and evaluated. For a typical multivariate model such as PLS regression, the reference properties are just column vectors, however, since deep learning modelling is based on utilising both the spatial and spectral information and aims to directly generate the prediction maps as output, the references for model training need to be supplied as expected prediction maps. To perform the DL, three separate ground-truth spatial maps of expected output were generated for segmentation, regression, and classification tasks. For the segmentation task, a binary image with background and manually segmented fruit image were used as the ground truth. For the regression task, each fruit was assigned the SSC values estimated using the refractometer. For the classification task, the walnut shells and kernel samples were manually labelled and used as the ground truth.

2.5. Data modelling

2.5.1. Image subsampling for deep learning

DL models require large data sets for training. A solution for deep learning on a small number of high-resolution spectral images was recently proposed as a sub-sampling of the spectral images to smaller size images [6,7,22]. In this study as well, to prepare the data set for the DL model, image subsampling was performed using the "extract_patches_2d" function from SciKit learn in Python. The images (PLS compressed) were sub-sampled into 500 sub-samples images of spatial dimension 256 × 256. With raw and SNV transformed data without any PLS compression, only 300 images were sub-sampled due to computer memory limitation for handling the

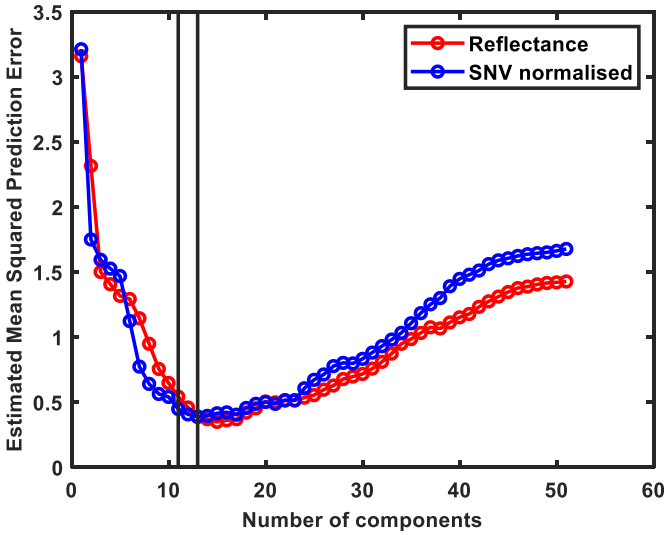


Fig. 3. Cross-validation plot for selection optimal LVs for PLS compression of reflectance and SNV normalised data.

Table 1
Jaccard score estimated for semantic segmentation performed using different forms of data.

Data forms	Jaccard score
Reflectance	0.98
SNV normalised reflectance	0.97
PLS compressed reflectance	0.99
PLS compressed SNV normalised	0.97

huge size spectral images. The trained models were tested on the independent test set of images which was never used for any model training and validation.

2.5.2. Deep learning with generative adversarial networks modelling

The aim of the generative model is to learn to transform an input image to the desired output form of the images, for example, transforming black and white images to color images, or transforming low-light images to bright light images, etc. The task of transforming images in different forms is termed image translation. In this study, it has been hypothesized that the GAN can be used to perform translation of the spectral images to output forms such as segmentation masks, property maps and classification maps. To perform image translation, a recently proposed cGANs approach was used [35]. The cGANs aims to learning the mapping from spectral images (or PLS transformed) x and random noise vector z to prediction maps y, G: $(x, z) \rightarrow y$, where G is the generator model trained to produce prediction maps that are like "real" images that will be screened by an adversarially trained discriminator, D. These prediction maps can be segmentation masks, fruit property maps and classification maps. Furthermore, unlike tradition pixel-wise chemometric analysis, the cGANs based modelling requires input and output as images. The G model is supplied with spectral images (or PLS transformed) to generate the prediction property maps. The D model uses both the spectral image (or PLS transformed) and the ground-truth prediction maps. The main aim of the D model is to detect if the G model synthesized prediction maps are of high quality or not. During the adversarial training process, the G model learns to generate high-quality prediction maps and eventually outperforms the D models' capability to detect differences between the ground truth and G synthesized prediction maps. Once the G model is trained it can be used to generate prediction maps depending on the required task (e.g., segmentation, regression, and classification). More details on the mathematical representation of the cGANs model are as follow.

The objective function of the cGANs [35] is in Eq. (1):

$$\mathscr{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[\log D(x,y))] + \mathbb{E}_{x,y}[\log(1 - D(x, G(x,z)))] \quad (1)$$

where G tries to minimize this objective function against an adversarial D that tries to maximize it [35] as Eq. (2).

$$G^* = \arg\min_G \max_D \mathscr{L}_{cGAN}(G,D) \quad (2)$$

To generate high-quality prediction maps, it is advised to combine the GAN with L1 loss [35]. The L1 on the G model is defined in Eq. (3):

$$\mathscr{L}_{L1}(G) = \mathbb{E}_{x,y,z}[y - G(x,z)_{\ 1}] \quad (3)$$

Combining (2) and (3) leads to final objective Eq. (4):

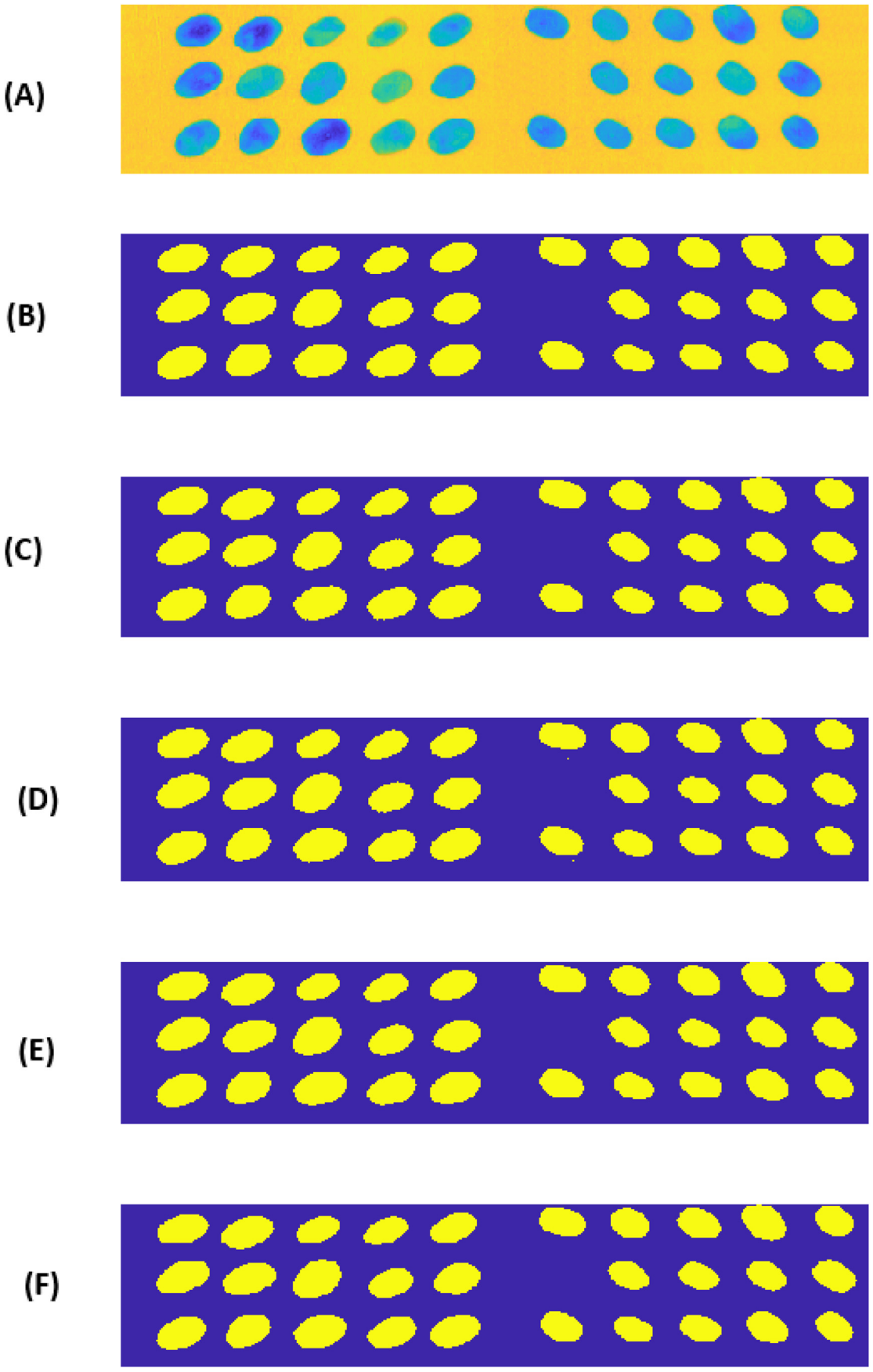
$$G^* = \arg\min_G \max_D \mathscr{L}_{cGAN}(G,D) + \lambda \mathscr{L}_{L1}(G) \quad (4)$$

The G model in this study was a "U-net" architecture [40] as explained in Fig. 2A, and the D model was a convolutional "PatchGAN" classifier as explained in Fig. 2B. The "PatchGAN" classifier was used to capture local style statistics [35]. At first, the G and D models have random weight, but during the adversarial training processed the G model weights are updated via L1 loss (mean absolute error) measured between the generated segmentation and the manually labelled segmentation maps. For D model, the final layer was fed to a sigmoid activation function for binary classification. The D model was compiled with the adaptive moment (ADAM) optimizer [41], with a learning rate, LR = 0.0002 and the loss function as 'binary_crossentropy' since the task was a binary classification of real or fake prediction maps. The 'binary_crossentropy' loss was also used to update the D model. Total epochs were 100, thus based on the training samples, the total number of iterations was 100 × number of samples. The model performance for segmentation and classification were judged based on Jaccard score [42] as can be understood in Eq. (5). The Jaccard score is defined as the size of the intersection divided by the size of the union of two label sets and used to compare predicted pixel labels to the corresponding set of ground truth labels. The Jaccard score range from 0 to 1, where Jaccard score of 0 explains no match, while Jaccard score of 1 mean perfect match between the predicted pixels labels and the corresponding ground truth labels. The performance for the regression task was judged based on the root mean squared error (RMSE) as defined in Eq. (6). In all cases, the performance of the model was also judged based on visual comparison with ground truth maps.

$$Jaccard\ score = \frac{|y \cap \hat{y}|}{|y \cup \hat{y}|} \quad (5)$$

$$RMSE(y,\hat{y}) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2} \quad (6)$$

The Jaccard score is estimated between two images i.e., prediction maps and ground truth maps, while the RMSE is used to judge the performance of the regression task that is based on the reference property and the mean predicted property value for each sample. DL model implementation and optimizations were done using the Python (3.6) language and the open-source deep learning


(A)
(B)
(C)
(D)
(E)
(F)

framework TensorFlow/Keras (2.4.0), running on a desktop workstation equipped with a NVidia GPU (GeForce RTX 2080 Ti), an Intel Xeon® W-2133 CPU @3.6 GHz and 64 GB RAM, running Microsoft Windows 10 OS. The manual data labelling, spectral pre-processing and PLS data compression was performed in MATLAB 2018b, Natick, MA, USA.

3. Results

The results for the segmentation, regression and classification analysis are presented in three separate sub-sections as follow.

3.1. Semantic segmentation analysis

The semantic segmentation was demonstrated using the grape data set where the aim was to segment the grape from the background. Furthermore, the semantic segmentation modelling was performed on four different forms of data i.e., raw reflectance, SNV normalised, PLS compressed reflectance data and PLS compressed SNV normalised data. For PLS compression, the optimal LVs were found using 10-fold cross-validation as presented in Fig. 3. For the reflectance data, a total of 13 LVs were chosen, and for the SNV normalised data, 11 LVs were chosen. The results of the semantic segmentation on the independent test set are shown in Table 1 and Fig. 4. It can be noted that for all the data forms, a very high Jaccard score was achieved (Table 1). A high Jaccard score i.e., close to 1 indicates ground truth resembling segmentation masks. The performance of semantic segmentation on PLS compressed reflectance data was marginally better than other forms of data. The semantic segmentation maps are shown in Fig. 4, furthermore, it can be noted that the predicted segmentation masks (Fig. 4C, D, E, F) were like the ground truth mask (Fig. 4B) demonstrating the potential of cGANs for semantic segmentation task. The main benefit of the PLS compression, in this case, was the reduction in the training time as for the complete spectral range the training time was 3 times more than the model training on PLS compressed data.

3.2. Semantic regression analysis

Like the semantic segmentation analysis, the cGANs was used for semantic regression where the SSC content in each grape was estimated. The semantic regression prediction maps are shown in Fig. 5, where most of the samples achieved similar SSC distribution maps (Fig. 5B, C, D, E) like the ground truth SSC maps (Fig. 5A). The performance of the prediction was further evaluated by extracting the mean predicted value for the grape and plotted against the reference SSC content as shown in Fig. 6. Note that the mean value for each fruit were obtained by averaging all the pixel predictions related to the fruit. It can be noted that the lowest RMSE was achieved with the PLS compressed reflectance data, followed by PLS compressed SNV normalised data. Like the semantic segmentation modelling, the semantic regression modelling benefitted with PLS compression both in lower RMSE and faster model training compared to using the full spectral range. In contrast, there was no added benefit of SNV normalization. As a comparison, pixel-wise analysis with PLS modelling was also performed and the example prediction maps are shown in Fig. 7C. Unlike the smooth prediction maps attained with cGANs modelling, the PLS based prediction maps were highly in-homogeneous. Such an in-homogeneity in prediction maps from pixel-based modelling is an inherent challenge of the pixel-wise analysis as it cannot handle the rich contextual information and is affected by minute pixel-to-pixel variations. Furthermore, in Fig. 7C it can be noted that the pixel-to-pixel variation is not random, but the prediction map has a distinct vertical pattern which could indicate that some spatial pixels in the camera might be having different sensitivity. On the other hand, the cGANs was able to avoid it and generated maps highly coherent like the ground truth mask.

3.3. Semantic classification

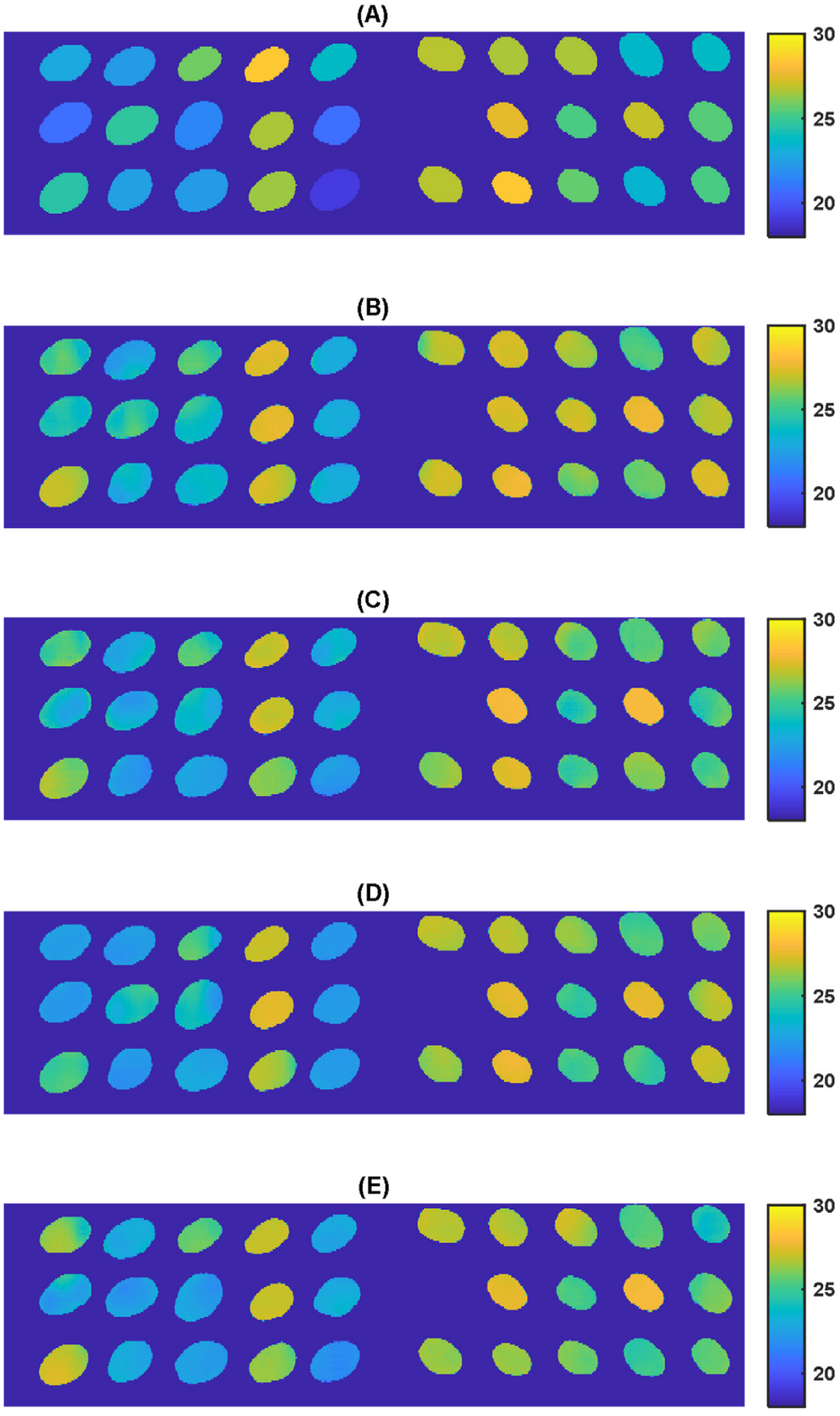
The cGANs was demonstrated for the classification of walnut shells and kernels. The case of kernels and shells classification was considered because spectral cameras are widely used in nuts processing lines for separating kernels from shells. At first, the spectral images were compressed using PLS loadings (Fig. 8), where the optimal loadings were identified with 10-fold cross-validation on walnut shells and kernels spectra on a dummy matrix. For reflectance, 5 LVs were selected and for SNV normalised data 3 LVs were selected for the compression. Later, four separate models were trained using the four different data forms and tested on the independent image test set. The Jaccard score for the prediction maps is shown in Table 2. It can be noted that the PLS compression did not contribute to achieving higher Jaccard scores as the scores were like the cGANs modelling performed on complete spectral range. However, the main benefit of the PLS compression was the faster model training (up to 3 times faster) compared to cGANs modelling on complete spectral range. The modelling performed on the SNV normalised spectral images were similar in performance on reflectance data. The Jaccard score is a measure of spatial similarity between the images, and it can be easily affected by differences in the detection of edges by the model. The Jaccard score is not related to the classification accuracy. This is noticeable that even with a Jaccard score of 0.60, the cGANs model was able to segment all the walnut shells and kernels and classify them correctly (Fig. 9) i.e., 100% correct classification at object level.

4. Discussion

The segmentation analysis presented in this study was based on the segmentation of the objects from the background and was particularly related to the segmentation of grapes from the background. Such a segmentation analysis in the traditional way of spectral image processing is the first step of the analysis which allows automatic detection of key objects of interest to extract the pixels for further modelling. However, a key point to note is that with cGANs modelling, the segmentation task can be directly integrated into the regression and classification tasks thus skipping the step of segmentation modelling. The integration of segmentation can be performed by providing the ground truth prediction maps for model training as pre-segmented maps like it was provided for the regression and classification case presented in this study (Figs. 5 and 9).

In the regression analysis presented in this study, the cGANs model was trained considering each pixel of the object to have the same average reference property (in the present study the average SSC of fruit juice). However, such a cGANs modelling is not feasible when the samples are highly heterogeneous, meaning, carrying spatial heterogeneity in the reference property. This is quite common, for example, in bigger fruit samples such as apple, pear and

Fig. 4. A summary of performance of cGANs models for semantic segmentation. (A) False colour spectral image, (B) ground truth segmentation mask, (C) segmentation mask with reflectance data, (D) segmentation mask with SNV normalised data, (E) segmentation mask with PLS compressed reflectance data, and (F) segmentation mask with PLS compressed SNV normalised data. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)


(A)
30
25
20
(B)
30
25
20
(C)
30
25
20
(D)
30
25
20
(E)
30
25
20

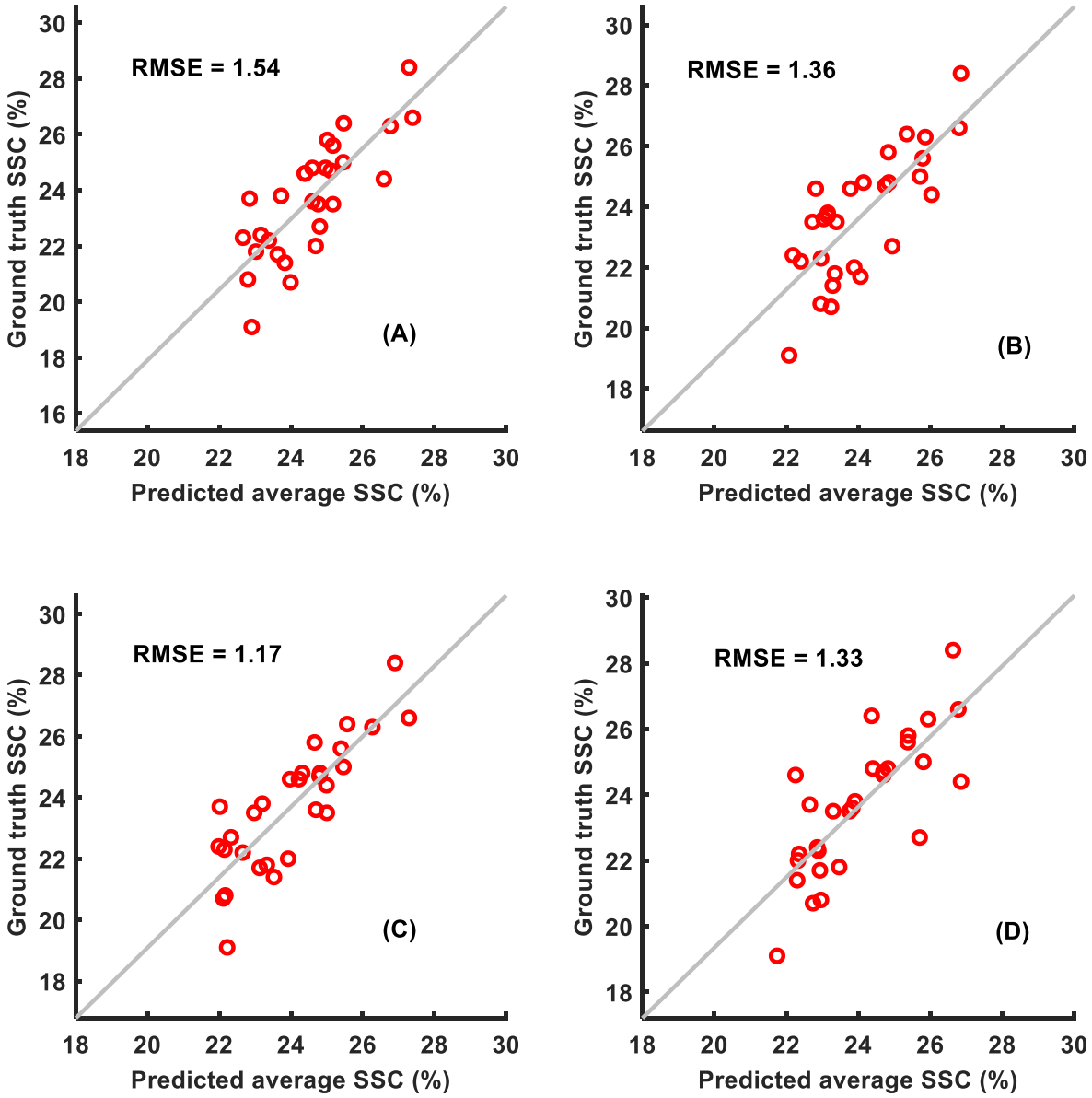


Fig. 6. Prediction plots for semantic regression to predict SSC in grapes. (A) Reflectance, (B) SNV normalised, (C) PLS compressed reflectance, and (D) PLS compressed SNV normalised data. The mean value for each fruit were obtained by averaging all the pixel predictions related to the fruit.

grape, where the SSC content varies based on the spatial location of fruit. A solution to that could be to perform multiple reference measurements on the same samples to also generate ground truth reference property maps capturing the spatial variability. However, this may be a challenging task to achieve in practice as it can take a lot of effort and time to perform sampling for reference analysis. Based on the analysis carried out in this study, cGANs based regression seems much straightforward for samples with homogeneous reference property compared to samples with spatial variation in reference property. But this limitation of cGANs is based on the challenge in obtaining ground truth reference property maps for heterogeneous samples and not on the modelling approach of cGANs. In the presence of heterogeneous reference property maps, the cGANs should perform equally well but needs to be verified in future works. Currently, for highly heterogenous samples the pixel-based modelling approaches should be sufficient. For classification modelling, the heterogeneity of the samples is not a problem as the cGANs model is trained using discrete class labels rather than continuous reference properties.

In this study for the regression modelling the cGANs model performed two tasks i.e., segmentation and reference property prediction, while for the classification case the cGANs model performed three tasks i.e., segmentation, classification of samples and separation of classification maps based on shells and kernels. The cGANs models can further be extended to performed operations such as bounding box generation and object counting tasks based on the need. However, the key message to note is that cGANs models are especially useful tools to perform multi-task operations.

Fig. 5. Semantic regression with cGANs for SSC prediction in grapes. (A) Ground truth SSC prediction maps, (B) Reflectance, (C) SNV normalised, (D) PLS compressed reflectance, and (E) PLS compressed SNV normalised data.

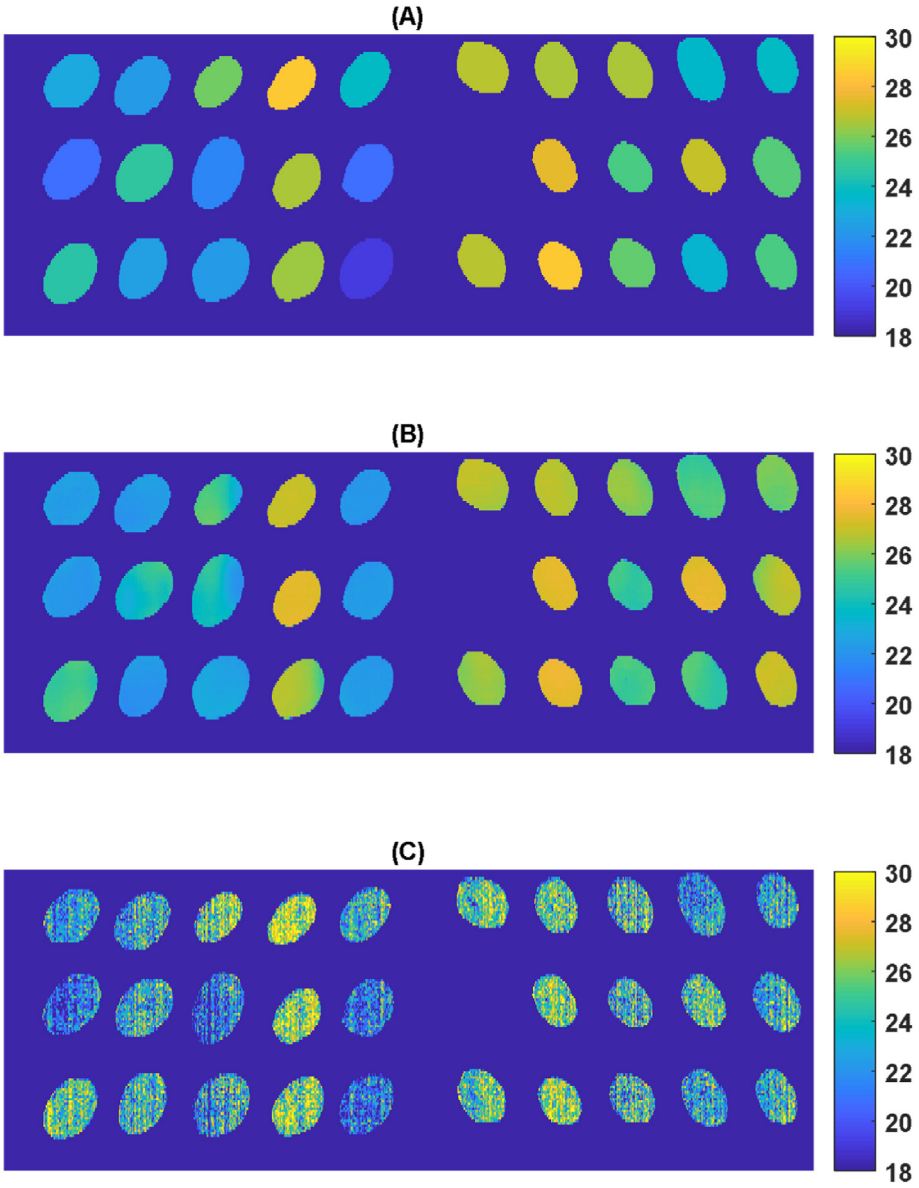


Fig. 7. A comparison of prediction maps attained with semantic regression and pixel-wise PLS regression-based analysis. (A) Ground truth SSC prediction maps, (B) Semantic regression on PLS compressed reflectance data, and (C) pixel-wise PLS regression analysis on reflectance data.

In the study, while comparing the prediction maps from the cGANs modelling with the pixel-based predictive modelling it was mentioned that the prediction maps were highly inhomogeneous in the spatial domain and carried some line patterns which can be related to the sensitivity of the sensor pixels. The spatial homogeneity in the grape can also be due to the spatial heterogeneity in reflectance property and the local curvature of the fruit surface. However, with the cGANs modelling, it was noted that the prediction maps were smoother and represented the average SSC content in the grape thus, bypassing both the line patterns and the effect of fruit surface curvature observed with pixel-based modelling. Such a better performance of the cGANs for generating homogenous prediction maps could be due to its capability to perform non-linear modelling as well as the contextual information present in the imaged scene. Nonetheless, further studies are needed to explore the potential of cGANs models to reduce and remove the heterogeneities due to sensor sensitivity and curvature of objects.

For deep learning modelling, it is important to have large data sets covering a wide range of sample variations, however, in the chemometric domain obtaining such data sets is a task a bit challenging due to the wet chemistry reference property analysis requirements. However, this limitation is related to regression modelling, as for classification modelling many images can be obtained. On other hands when the images are of sufficiently high resolution and contain several objects in the scene, the image augmentation approach based on image sub-sampling can be used for model training. In this study, the image sub-sampling approach presented in earlier studies [6,7,22] allowed sufficient sub-samples for training deep learning models from only one high-resolution spectral image. However, a key point to note is that real data cannot replace the augmented data and whenever possible it is always better to acquire larger data sets. Image augmentation can also be performed in other ways such as image rotation, horizontal and vertical shifts, shear, and flips. Nonetheless, in this study, it was

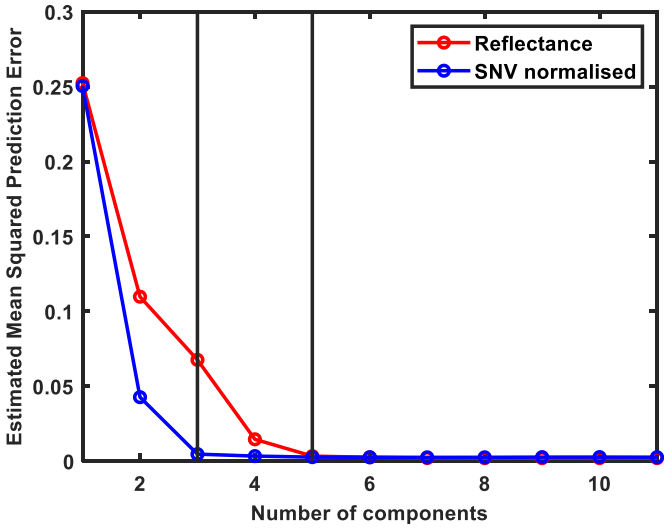


Fig. 8. Cross-validation plot for selection optimal LVs for PLS compression of reflectance and SNV normalised data.

Table 2
Quality of semantic classification maps attained with modelling performed on difference forms of data.

Data forms	Jaccard score
Reflectance	0.59 ± 0.09
SNV normalised reflectance	0.60 ± 0.09
PLS compressed reflectance	0.59 ± 0.10
PLS compressed SNV normalised	0.54 ± 0.07

not in the scope as the model achieved sufficient capabilities to perform the segmentation, regression, and classification. In future work, the effect of different augmentation approaches could be a direction to explore to achieve robust and generalized models.

In earlier studies related to deep learning modelling in the chemometric literature, the effect of spectral pre-processing and data compression with latent variable approaches such as principal component analysis (PCA) have shown to achieve improved models, particularly related to the improvement of model training with spectral normalization and reduction in time requirements with data compression [6,7]. Hence, in this study, the effect of pre-processing data with SNV normalization and the PLS based data compression was explored. The SNV normalization was used as it can suppress the illumination heterogeneities due to the local curvature on objects [10]. Like earlier studies, in this study it was found that the data compression directly reduced the model training time (up to 3 times less) compared to modelling performed on the full spectral range, however, the SNV normalization did not show any improvement in model accuracies. On the other hand, the PLS compression before the cGANs modelling led to lower RMSE models indicating that apart from a reduction in training time, the PLS compression complemented the regression modelling.

The cGANs based approach presented in this work requires an image as input to translate it to output. This scenario is perfectly suited for the new snapshot spectral cameras as they directly generate the complete spectral image in one capture. On other hand, the most popular spectral cameras currently used in practice are line scan cameras which acquires data line by line. To use the cGANs based approach in real-time for line scan cameras the user need to either wait for the complete image to be acquired (in case the experiment involves discrete samples) or for in-line application wait for the initial $k-1$ lines to have been acquired for a cGANs model of $k \times k$ spatial resolution and later using the cGANs model on the image generated during every new line acquisition similar to as proposed in Ref. [43].

One of the key benefits of the traditional chemometric approaches is that they are highly parsimonious and particularly key information such as scores, loading and regression vectors can provide detailed insights to the underlying physicochemical information being modelling. The DL based models are less parsimonious in terms of gaining detailed insights to the background physicochemical information. However, in recent years, progress is being made towards explainable deep learning and different concepts such as Grad-CAM, integrated gradients, are emerging. In the chemometrics domain, recently the implementation of Grad-CAM [44] for spectral data modelling has shown insights into the key spectral variables contributing to the model. Although, the research towards an explanation of the DL model is still in the very early stage, and in coming years, better approaches to understanding DL models should become available. Implementing an explainable DL method for hyperspectral imaging will be a direction for future work.

5. Conclusions

This study presented an innovative approach to model spectral imaging data using an artificial intelligence approach based on conditional generative adversarial networks. This novel approach treated the spectral images as images and used both the spatial and spectral information with convolutional operations to perform three main tasks for spectral image processing i.e., segmentation, regression, and classification. The result showed that this novel approach achieved much smoother prediction maps compared to achievable with pixel-based modelling. Furthermore, the cGANs models were able to perform multiple tasks, for example, for regression analysis, the cGANs model directly performed the segmentation and prediction of reference property. Similarly, for the classification case, multiple tasks such as segmentation, classification of samples and separation of classification maps for individual objects, were performed in a single model. The ability to cGANs model to directly take the spectral image as input and provide output as prediction maps allows it to bypass several steps such as spectra image unfolding and refolding, region of interest selection for selecting spectra for chemometric modelling and pixel-wise modelling and application. Furthermore, it was found that the PLS based data compression can benefit the cGANs models by reducing the model training time and achieving models with lower prediction error. A key point to note is that in the chemometric domain, one of the main interests is to know the background chemistry of samples with the use of traditional loading and regression vectors, however, such enhanced insights cannot be achieved with cGANs modelling. Therefore, the cGANs can be useful for practical implementation or to be integrated into easy-to-use automated black box software's for spectral image processing, as currently the capability for inferring information based on cGANs about the background chemistry is limited. To understand the background chemistry of samples in detail, readers should

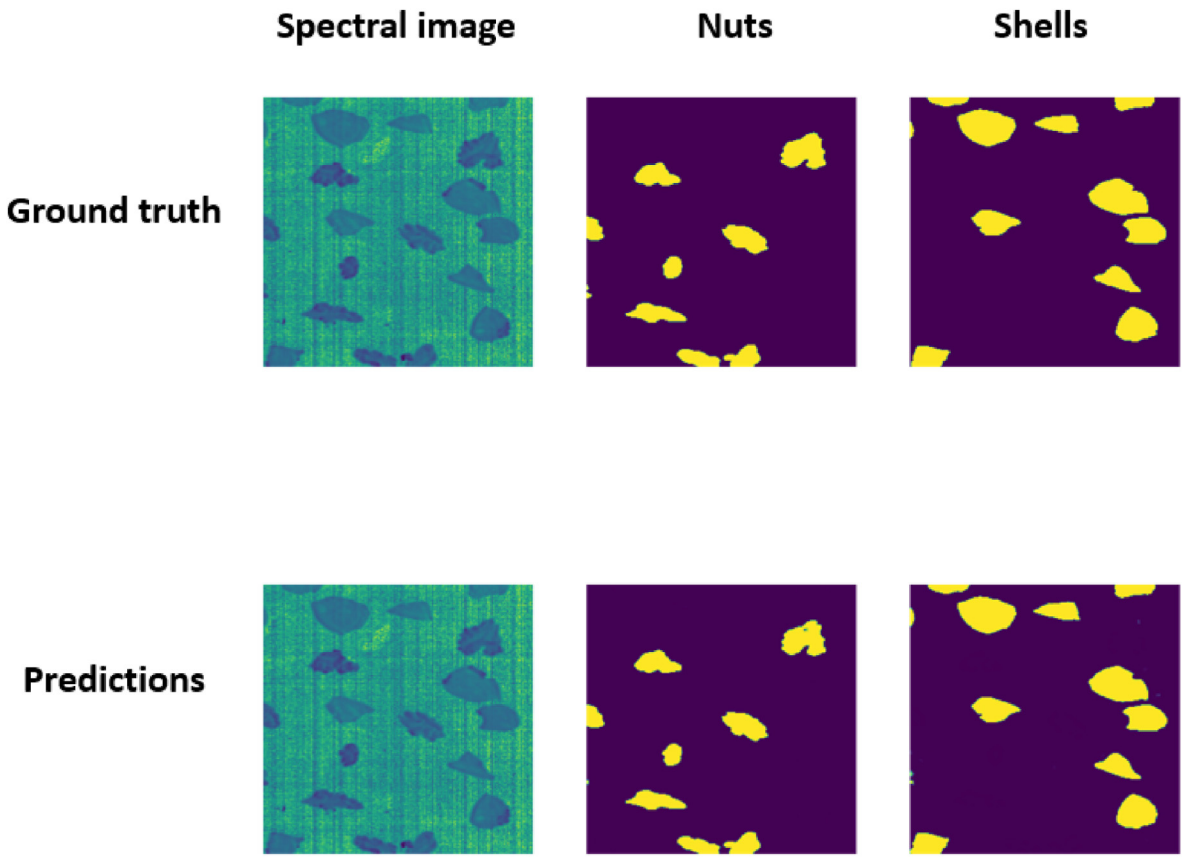


Fig. 9. An example of the cGANs model for classification of walnut shells and nuts. In top row the ground truth classification masks are shown while in bottom row the cGANs synthesized classification maps are shown. It can be noted that although the Jaccard score was 0.60, however, all the nuts and shells were classified into correct classes.

implement the traditional chemometric approaches to spectral image processing. Although, some recent deep learning approaches such as GradCAM can provide insight into key spectral bands and spatial features related to the model predictive power.

CRediT authorship contribution statement

Puneet Mishra: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft.

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 author would like to express thanks to Dário Passos for his valuable comments and suggestions during the development of the manuscript.

References

[1] J.M. Amigo, I. Martí, A. Gowen, F. Marini, Chapter 9 - Hyperspectral Imaging and Chemometrics: A Perfect Combination for the Analysis of Food Structure, Composition and Quality, Data Handling in Science and Technology, Elsevier, 2013, pp. 343–370.
[2] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, J.M. Frias, Hyperspectral imaging – an emerging process analytical tool for food quality and safety control, Trends Food Sci. Technol. 18 (2007) 590–598.
[3] J.M. Amigo, H. Babamoradi, S. Elcoroaristizabal, Hyperspectral image analysis. A tutorial, Anal. Chim. Acta 896 (2015) 34–51.
[4] N. Mobaraki, J.M. Amigo, HYPER-Tools. A graphical user-friendly interface for hyperspectral image analysis, Chemometr. Intell. Lab. Syst. 172 (2018) 174–187.
[5] M. Vidal, J.M. Amigo, Pre-processing of hyperspectral images. Essential steps before image analysis, Chemometr. Intell. Lab. Syst. 117 (2012) 138–148.
[6] P. Mishra, R. Sadeh, E. Bino, G. Polder, M.P. Boer, D.N. Rutledge, I. Herrmann, 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.
[7] P. Mishra, R. Sadeh, M. Ryckewaert, E. Bino, G. Polder, M.P. Boer, D.N. Rutledge, I. Herrmann, 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.
[8] J. Bøtker, J.X. Wu, J. Rantanen, J.M. Amigo, Chapter 3.7 - Hyperspectral Imaging as a Part of Pharmaceutical Product Design, Data Handling in Science and Technology, Elsevier2020, pp. 567-581.
[9] A. Brugger, J. Behmann, S. Paulus, H.-G. Luigs, M.T. Kuska, P. Schramowski, K. Kersting, U. Steiner, A.-K. Mahlein, Extending hyperspectral imaging for plant phenotyping to the UV-range, Rem. Sens. 12 (11) (2019).
[10] P. Mishra, S. Lohumi, H. Ahmad Khan, A. Nordon, Close-range hyperspectral imaging of whole plants for digital phenotyping: recent applications and illumination correction approaches, Comput. Electron. Agric. 178 (2020) 105780.
[11] P. Mishra, M.S.M. Asaari, A. Herrero-Langreo, S. Lohumi, B. Diezma, P. Scheunders, Close range hyperspectral imaging of plants: a review, Biosyst. Eng. 164 (2017) 49–67.
[12] P. Lasch, M. Stämmler, M. Zhang, M. Baranska, A. Bosch, K. Majzner, FT-IR hyperspectral imaging and artificial neural network analysis for identification of pathogenic bacteria, Anal. Chem. 90 (2018) 8896–8904.
[13] A.A. Gowen, C. O'Sullivan, C.P. O'Donnell, Terahertz time domain spectroscopy and imaging: emerging techniques for food process monitoring and quality control, Trends Food Sci. Technol. 25 (2012) 40–46.
[14] S. Lohumi, M.S. Kim, J. Qin, B.-K. Cho, Raman imaging from microscopy to macroscopy: quality and safety control of biological materials, Trac. Trends Anal. Chem. 93 (2017) 183–198.
[15] L. Coic, P.-Y. Sacré, A. Dispas, C. De Bleye, M. Fillet, C. Ruckebusch, P. Hubert, E. Ziemons, Pixel-based Raman hyperspectral identification of complex

pharmaceutical formulations, Anal. Chim. Acta 1155 (2021) 338361.
[16] G. Wan, G. Liu, J. He, R. Luo, L. Cheng, C. Ma, Feature wavelength selection and model development for rapid determination of myoglobin content in nitrite-cured mutton using hyperspectral imaging, J. Food Eng. 287 (2020) 110090.
[17] J. Ma, D.-W. Sun, Prediction of monounsaturated and polyunsaturated fatty acids of various processed pork meats using improved hyperspectral imaging technique, Food Chem. 321 (2020) 126695.
[18] P.T. Sorenson, S.A. Quideau, B. Rivard, M. Dyck, Distribution mapping of soil profile carbon and nitrogen with laboratory imaging spectroscopy, Geoderma 359 (2020) 113982.
[19] H. Zhang, B. Zhan, F. Pan, W. Luo, Determination of soluble solids content in oranges using visible and near infrared full transmittance hyperspectral imaging with comparative analysis of models, Postharvest Biol. Technol. 163 (2020) 111148.
[20] Y. Dixit, M. Al-Sarayreh, C.R. Craigie, M.M. Reis, A Global Calibration Model for Prediction of Intramuscular Fat and pH in Red Meat Using Hyperspectral Imaging, Meat Science, 2020, p. 108405.
[21] M. Mäkelä, P. Geladi, M. Rissanen, L. Rautkari, O. Dahl, Hyperspectral near infrared image calibration and regression, Anal. Chim. Acta 1105 (2020) 56–63.
[22] P. Mishra, I. Herrmann, GAN Meets Chemometrics: Segmenting Spectral Images with Pixel2pixel Image Translation with Conditional Generative Adversarial Networks, Chemometrics and Intelligent Laboratory Systems, 2021.
[23] J.-L. Xu, S. Hugelier, H. Zhu, A.A. Gowen, Deep learning for classification of time series spectral images using combined multi-temporal and spectral features, Anal. Chim. Acta 1143 (2021) 9–20.
[24] J.-L. Xu, C. Riccioli, A. Herrero-Langreo, A.A. Gowen, Deep learning classifiers for near infrared spectral imaging: a tutorial, J. Spectr. Imaging 9 (2020).
[25] A. Herrero-Langreo, N. Gorretta, B. Tisseyre, A. Gowen, J.-L. Xu, G. Chaix, J.-M. Roger, Using spatial information for evaluating the quality of prediction maps from hyperspectral images: a geostatistical approach, Anal. Chim. Acta 1077 (2019) 116–128.
[26] N. Gorretta, J. Roger, G. Rabatel, V. Bellon-Maurel, C. Fiorio, C. Lelong, Hyperspectral Image Segmentation: the Butterfly Approach, 2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009, pp. 1–4.
[27] A. Nardecchia, R. Vitale, L. Duponchel, Fusing spectral and spatial information with 2-D stationary wavelet transform (SWT 2-D) for a deeper exploration of spectroscopic images, Talanta 224 (2021) 121835.
[28] M. Ahmad, R. Vitale, C.S. Silva, C. Ruckebusch, M. Cocchi, Exploring local spatial features in hyperspectral images, J. Chemometr. 34 (2020), e3295.
[29] P. Mishra, A. Karami, A. Nordon, D.N. Rutledge, J.M. Roger, Automatic de-noising of close-range hyperspectral images with a wavelength-specific shearlet-based image noise reduction method, Sensor. Actuator. B Chem. 281 (2019) 1034–1044.
[30] P. Mishra, A. Nordon, M.S. Mohd Asaari, G. Lian, S. Redfern, Fusing spectral and textural information in near-infrared hyperspectral imaging to improve green tea classification modelling, J. Food Eng. 249 (2019) 40–47.
[31] J. Chai, H. Zeng, A. Li, E.W.T. Ngai, Deep learning in computer vision: a critical review of emerging techniques and application scenarios, Machine Learn. Appl. 6 (2021) 100134.
[32] L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, L. Farhan, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, J. Big Data 8 (2021) 53.
[33] N. Audebert, B.L. Saux, S. Lefevre, Deep learning for classification of hyperspectral data: a comparative review, IEEE Geosci. Remote Sens. Magazine 7 (2019) 159–173.
[34] M.E. Paoletti, J.M. Haut, J. Plaza, A. Plaza, Deep learning classifiers for hyperspectral imaging: a review, ISPRS J. Photogrammetry Remote Sens. 158 (2019) 279–317.
[35] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image Translation with Conditional Adversarial Networks, pp. 1125-1134.
[36] R. Soni, T. Arora, A. Solanki, A. Nayyar, M. Naved, Chapter 5 - A Review of the Techniques of Images Using GAN, Generative Adversarial Networks for Image-To-Image Translation, Academic Press2021, pp. 99-123.
[37] P. Mishra, M. Sytsma, A. Chauhan, G. Polder, E. Pekkeriet, All-in-one: a spectral imaging laboratory system for standardised automated image acquisition and real-time spectral model deployment, Anal. Chim. Acta 1190 (2022) 339235.
[38] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, Anal. Chem. 36 (1964) 1627–1639.
[39] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777.
[40] O. Ronneberger, P. Fischer, T. Brox, U-Net, Convolutional Networks for Biomedical Image Segmentation, Springer International Publishing, Cham, 2015, pp. 234–241.
[41] D.P. Kingma, J. Ba, Adam, A Method for Stochastic Optimization, arXiv preprint arXiv:1412.6980, 2014.
[42] T.T. Tanimoto, Elementary Mathematical Theory of Classification and Prediction, 1958.
[43] R. Rocha de Oliveira, A. de Juan, SWiVIA – sliding window variographic image analysis for real-time assessment of heterogeneity indices in blending processes monitored with hyperspectral imaging, Anal. Chim. Acta 1180 (2021) 338852.
[44] 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.