Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy xxx (xxxx) 124579

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

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

Differentiation of *Listeria monocytogenes* serotypes using near infrared hyperspectral imaging

Rumbidzai T. Matenda [a], Diane Rip [a], J.A. Fernández Pierna [b], Vincent Baeten [b], Paul J. Williams [a, *]

[a] Department of Food Science, Stellenbosch University, Private Bag X1, Matieland, Stellenbosch 7602, South Africa
[b] Quality and authentication of products Unit, Knowledge and valorization of agricultural products Department, Walloon Agricultural Research Centre (CRA-W), Chaussée de Namur,24, 5030 Gembloux, Belgium

ARTICLE INFO

Keywords:
Food pathogens
Listeria monocytogenes serotypes
NIR-HSI
Multivariate data analysis
Partial least discriminant analysis
Principal component analysis

ABSTRACT

Among the severe foodborne illnesses, listeriosis resulting from the pathogen *Listeria monocytogenes* exhibits one of the highest fatality rates. This study investigated the application of near infrared hyperspectral imaging (NIR-HSI) for the classification of three *L. monocytogenes* serotypes namely serotype 4b, 1/2a and 1/2c. The bacteria were cultured on Brain Heart Infusion agar, and NIR hyperspectral images were captured in the spectral range 900–2500 nm. Different pre-processing methods were applied to the raw spectra and principal component analysis was used for data exploration. Classification was achieved with partial least squares discriminant analysis (PLS-DA). The PLS-DA results revealed classification accuracies exceeding 80 % for all the bacterial serotypes for both training and test set data. Based on validation data, sensitivity values for *L. monocytogenes* serotype 4b, 1/2a and 1/2c were 0.69, 0.80 and 0.98, respectively when using full wavelength data. The reduced wavelength model had sensitivity values of 0.65, 0.85 and 0.98 for serotype 4b, 1/2a and 1/2c, respectively. The most relevant bands for serotype discrimination were identified to be around 1490 nm and 1580–1690 nm based on both principal component loadings and variable importance in projection scores. The outcomes of this study demonstrate the feasibility of utilizing NIR-HSI for detecting and classifying *L. monocytogenes* serotypes on growth media.

1. Introduction

The World Health Organization reports that bacterial pathogens account for over 30 % of all foodborne illnesses globally [1]. In Africa alone, approximately 92 million people suffer from foodborne illnesses resulting in nearly 137 000 deaths annually [2,3]. *Listeria monocytogenes* is one example of pathogenic bacteria commonly found in ready-to-eat foods such as cooked ham, polony, vegetables, soft cheese, and raw milk [4,5]. Cases resulting from *L. monocytogenes* infections are low as compared to other pathogenic bacteria such as *Escherichia coli* and *Salmonella* spp. However, the mortality rate of listeriosis (the disease caused by *L. monocytogenes)* can reach up to 30 %, with the elderly, pregnant individuals and immune compromised people being primarily at risk [6].

L. monocytogenes, a rod shaped, Gram-positive bacterial pathogen is one of seventeen species within the genus *Listeria*. According to Orsi and Wiedmann [7], *L. monocytogenes* has thirteen serotypes which are classified into lineages I-IV. Lineage I isolates, specifically serotypes 1/2b and 4b have been attributed to most human cases of listeriosis [8,9]. Nevertheless, isolates from lineage II (serotypes 1/2a and 1/2c) also play a role in listeriosis, with serotype 1/2a being more prevalent in hospitalisation cases [10]. According to Poimenidou *et al.* [11], serotype 1/2a strains are usually isolated from food and food environments whilst serotype 4b strains arises from various sources including soil and water. Previous research has proven that under appropriate environmental conditions, it is possible for *L. monocytogenes* serotypes 1/2a, 1/2b, 1/2c, and 4b to coexist in food factories [12]. For example, a study on fresh seafood samples across supermarkets in Iran found that serotypes 1/2a, 1/2b, and 4b were all present in fish and shrimp samples investigated [13]. The bacterium can survive extreme temperatures (1–45 °C) and wide pH ranges (4.5–9.0), making it a robust food pathogen [14]. Thus, it is important to detect this bacterial contaminant using fast and efficient methods to prevent outbreaks and product recalls.

Various methods have been developed over the years, with the aim of detecting specific food pathogens more rapidly and accurately. The

* Corresponding author.
E-mail address: pauljw@sun.ac.za (P.J. Williams).

https://doi.org/10.1016/j.saa.2024.124579
Received 29 February 2024; Received in revised form 27 May 2024; Accepted 30 May 2024
1386-1425/

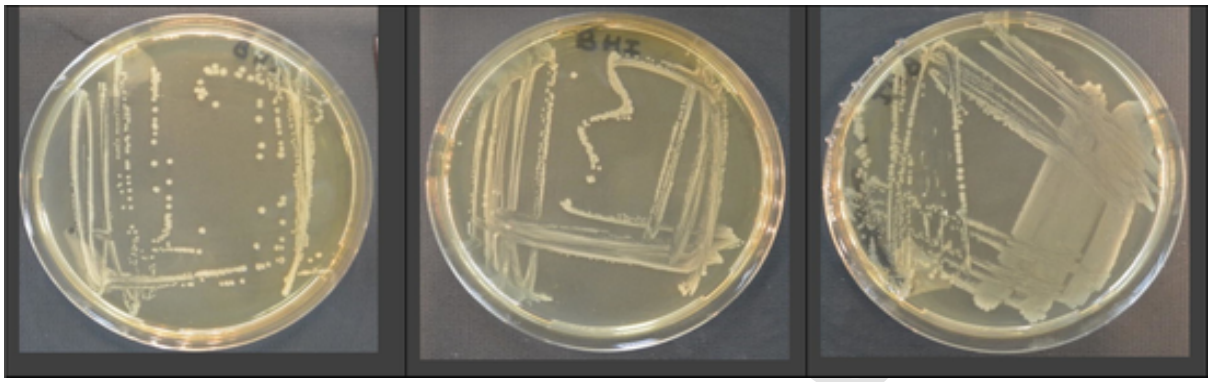

Fig. 1. Digital images of *Listeria* serotypes streaked onto BHI agar (from left) *L. monocytogenes* 4b, *L. monocytogenes* 1/2a and *L. monocytogenes* 1/2c (Samsung A31; 1080 x 2400 pixels).

most common types of methods include nucleic acid-based and immunological based methods [15]. Both polymerase chain reaction (PCR) and enzyme-linked immuno-sorbent assay (ELISA) are good examples of methods proven to improve accuracy and the time it takes for a laboratory to process a sample and provide the results (turn-around time). However, these methods still require at least two hours before results can be obtained which makes them unsuitable for automation [16]. Considering such limitations, it is highly desirable to develop non-contact, rapid methods for bacterial analysis.

The use of visible (VIS) and near infrared (NIR) spectroscopy has also been previously explored for bacterial analysis with high success [17,18]. Feng *et al.* [19], investigated the use of visible/NIR spectroscopic data to detect and differentiate between various *E. coli* and *L. innocua* strains. The results showed that the application of least squares support vector machines (LS-SVM) resulted in classification accuracies of above 85 %. In addition, due to improvements and advancements in technology, the use of other rapid methods of analysis like hyperspectral imaging (HSI) have been investigated. HSI is a robust technique which has proven to be successful in the detection and classification of pathogens on various foods and nutrient rich media [20–22]. The technique combines spatial and spectral data, providing detailed spectral information for each captured pixel. By analysing the spectral data collected through NIR-HSI, accurate classification and identification of multiple bacteria species within a single food sample can be achieved. It has been proven that, when using HSI, bacterial classification is based on differences in the cell wall's chemical composition of the different bacterial species and serogroups at selective wavelengths [23,24]. For instance, in one study, differences in protein content were used to distinguish between two lactic acid bacterial species in cooked ham samples [25]. The technology is particularly valuable for tasks that require fine discrimination between materials based on their spectral characteristics.

Yoon *et al.* [26], investigated the use of visible/ near infrared (400–1000 nm) hyperspectral imaging to identify six different *E. coli* serotypes on rainbow agar using principal component analysis, Mahalanobis distance (PCA-MD) and PCA, k-nearest neighbours (PCA-KNN). Both approaches achieved a classification accuracy of above 84 % for all serogroups. However, it is important to note the impact of data pre-processing. Different pre-processing techniques were employed, and the results showed that models pre-processed with SNVD (standard normal variant and detrending) were superior compared to others investigated. In another study, researchers used HSI in the NIR region (1000–2500 nm) to classify *Staphylococcus aureus* and *S. epidermidis* at strain level on solid Luria-Bertani agar [20]. The authors used partial least discriminant analysis (PLS-DA) and results showed classification accuracies of above 90 % despite the spectral similarities.

Advancements in chemometrics has also led to substantial enhancements in predictive modelling through the utilization of variable selection algorithms and combined classification algorithms. A study by Feng *et al.* [24] used NIR-HSI for classification of different bacterial species on the same media (i.e., Tryptone Soy Agar) in the wavelength range 400–1000 nm. PLS-DA was compared to an optimized support vector machine (SVM) model. The SVM model was optimized with the invasive weed algorithm to yield invasive weed optimization support vector machines (IWO-SVM) aimed at improving the model's accuracy. Various variable selection methods such as genetic algorithm (GA), successive projection algorithm (SPA) and competitive adaptive reweighted sampling (CARS) were used to select important spectral bands. Despite the different variable selection techniques, the results illustrated that the PLS-DA models performed poorly when compared to the SVM models. The overall correct classification accuracies (OCCRs) were 41 % and 91 % for PLS-DA and IWO-SVM, respectively. This indicated that PLS-DA (in that study) was not suitable for the classification of the bacterial strains investigated. Gu *et al.* [27] adopted a more universal approach with the specific aim of investigating the feasibility of distinguishing between pathogenic bacteria, including *E. coli, S. aureus*, and *Salmonella,* cultured on different agar media, using a single model approach. Three wavelength selection techniques were used in conjunction with PLS-DA and optimized SVM (Grasshopper optimisation algorithm-SVM (GOA-SVM)) algorithms. Results also demonstrated poor performances of OCCRs for PLS-DA models suggesting that linear models might not be suitable when dealing with cases of closely related bacteria. However, GOA-SVM performed well with OCCRs of above 98 % for both calibration and prediction sets.

The aforementioned studies show that NIR-HSI together with chemometric techniques is successful in detecting and predicting bacteria species/serotypes. This functionality positions NIR-HSI as a promising tool for enhancing food safety monitoring and quality control in the food industry, as it enables rapid and reliable detection of potential microbial contaminants. However, few have investigated bacteria from the same species. Hence, the aim of this study was to investigate the application of NIR-HSI for the potential prediction of three *L. monocytogenes* serotypes 4b, 1/2a and 1/2c on solid media.

2. Materials and methods

2.1. Bacterial culture and sample preparation

Three *L. monocytogenes* serotypes were investigated, including one from lineage I (*L. monocytogenes* 4b (ATCC 23074) and two from lineage II (*L. monocytogenes* 1/2c (ATCC 7644) and *L. monocytogenes* 1/2a (ATCC 19111)). All bacterial cultures were obtained in lyophilized form from Davies Diagnostics, South Africa, and resuspended as per the man-

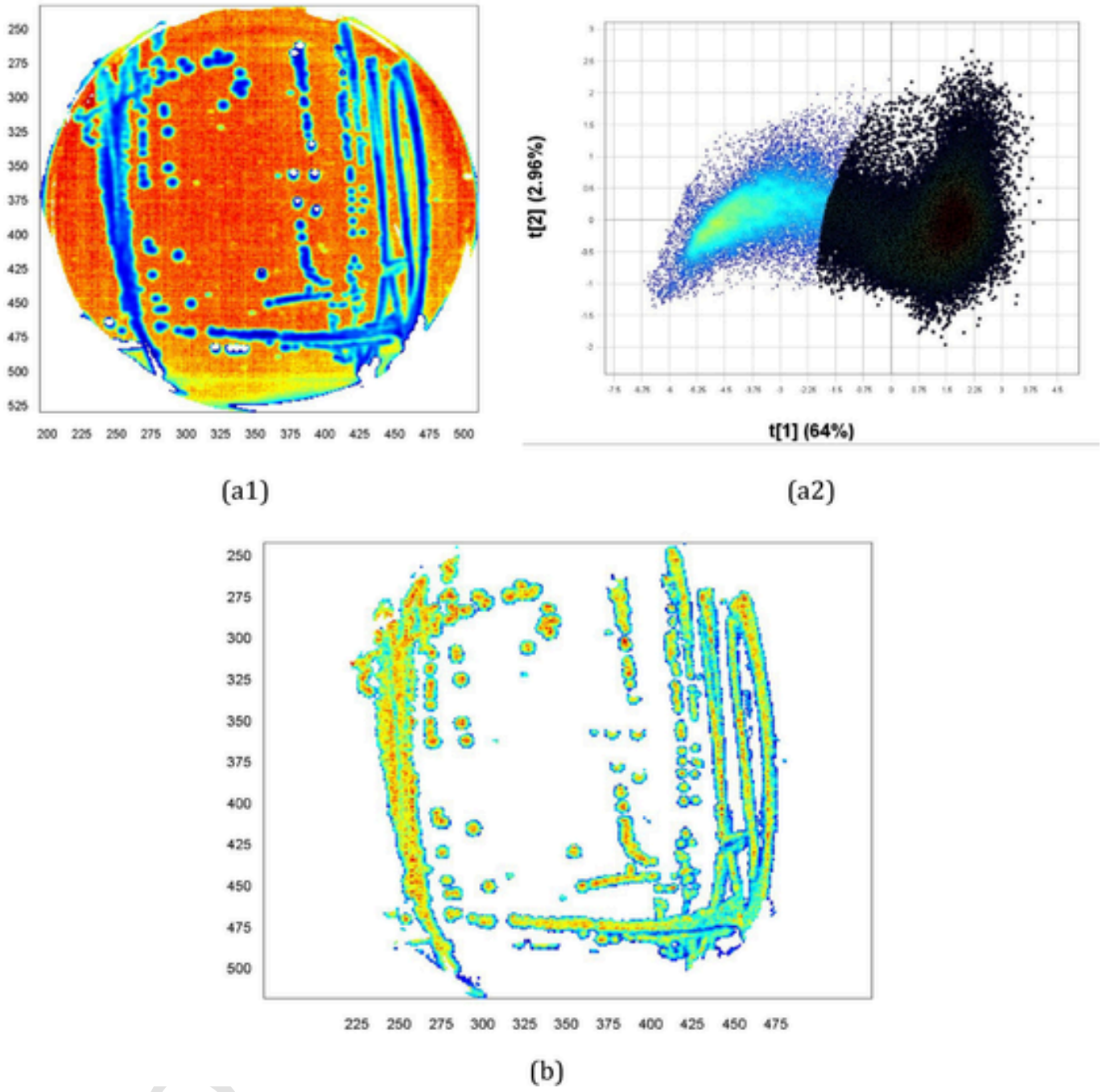

Fig. 2. (a1&a2) PC1 score image of *L. monocytogenes* colonies and BHI agar and the corresponding principal component analysis score plot showing PC1 and PC2. These were used iteratively to select the regions of interest (i.e., the bacteria). (b) The resultant PC1 score image of *L. monocytogenes.*

Table 1
Calibration and validation set sizes for pixel-wise analysis based on the number of spectra.

	L. monocytogenes 4b	*L. monocytogenes* 1/2a	*L. monocytogenes* 1/2c	Total
Calibration	3432	3074	3148	9654
Validation	2525	2749	2416	7690
Total	5957	5823	5564	17,344

ufacturer's instructions. Cultures were stored in skim milk tryptone glucose glycerin (STGG) tubes (National Health Laboratory Service, Greenpoint) at −80 ^{0}C until needed. Brain Heart Infusion (BHI) (Oxoid, United Kingdom) agar (a general-purpose growth medium, specific for fastidious organisms) was used throughout the study for uniformity and minimization of spectral variation from the growth media. To account for colony concentration variations, the streaking method was used, which allowed for single colonies as well as larger areas of confluent growth.

From STGG, a loopful (˃1 μl) of individual stock culture was streaked onto BHI agar under aseptic conditions in a class II biosafety cabinet. For optimum microbial growth, the petri dishes/plates were

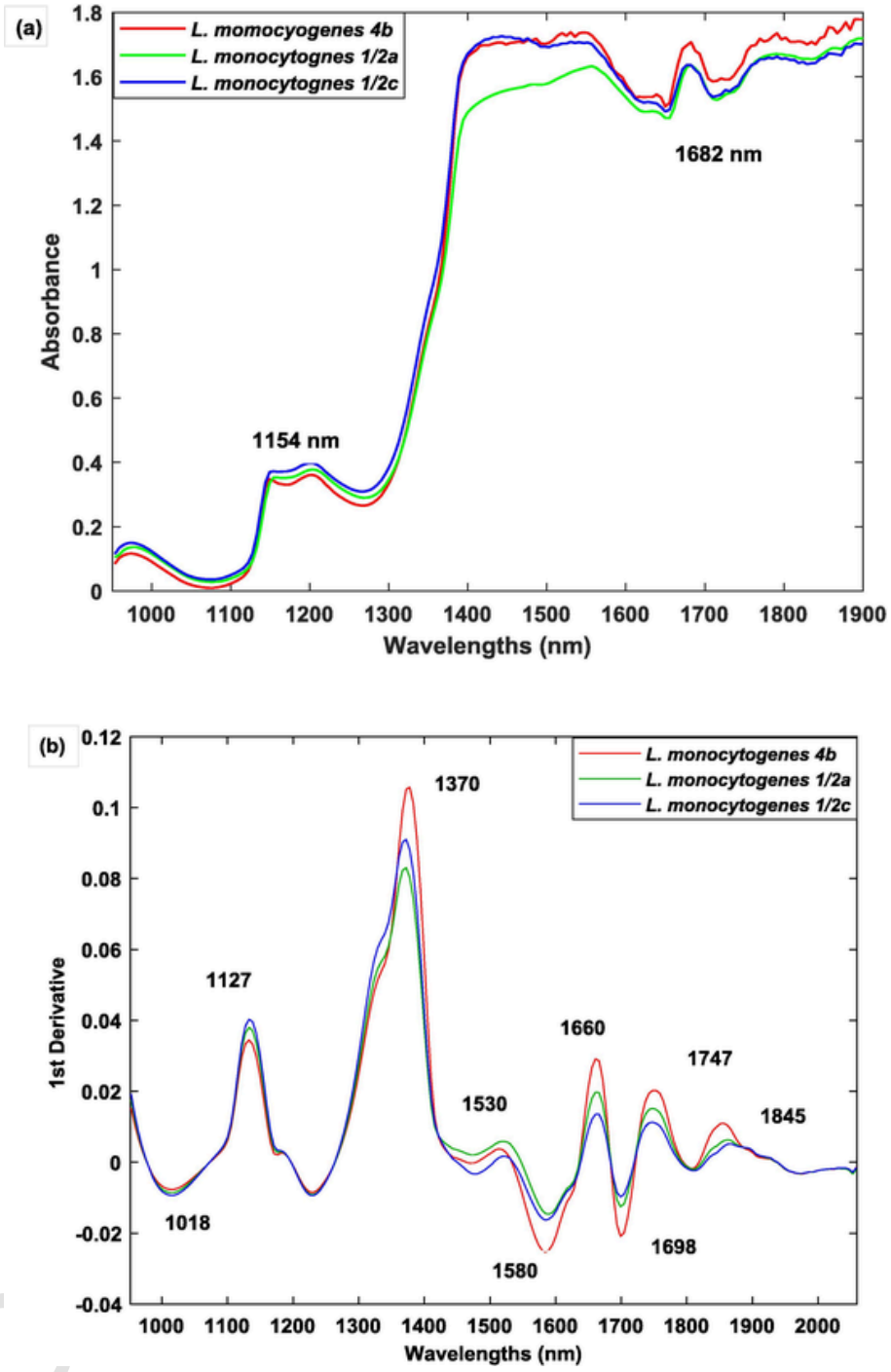

Fig. 3. (a) The original raw absorbance spectra of bacterial colonies of *L. monocytogenes* serotype 4b (red), serotype 1/2a (green) and serotype 1/2c (blue). (b) Savitzky Golay (1st derivative 2nd polynomial 11-point smoothing) pre-processed spectra. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

thereafter incubated at 37 ^{0}C for 24 h. As a precaution, this process was repeated to ensure the viability and purity of the bacteria [28]. Thereafter, petri dishes were prepared in duplicate for each *L. monocytogenes* serotype by taking a single colony of the bacteria from the incubated plates. The bacteria were streaked onto BHI and incubated for 22 ± 1 h at 37 ^{0}C (Fig. 1). Prior to hyperspectral imaging, the plates were allowed to cool down to ambient temperature (approximately 21 ^{0}C) for 15 min. This procedure was considered as the first experiment (Rep 1). The protocol was repeated, resulting in replicate samples (Rep 2) and an overall total of 12 petri dishes.

2.2. NIR-HSI imaging system and image acquisition

NIR hyperspectral images were acquired in an air-conditioned room set at 21 ^{0}C using a line scan HySpex SWIR 384 (short wave infrared) camera (Norsk Elektro Optikk (NEO), Norway). The HSI system comprised of a spectrograph, translation stage, a mercury-cadmium-

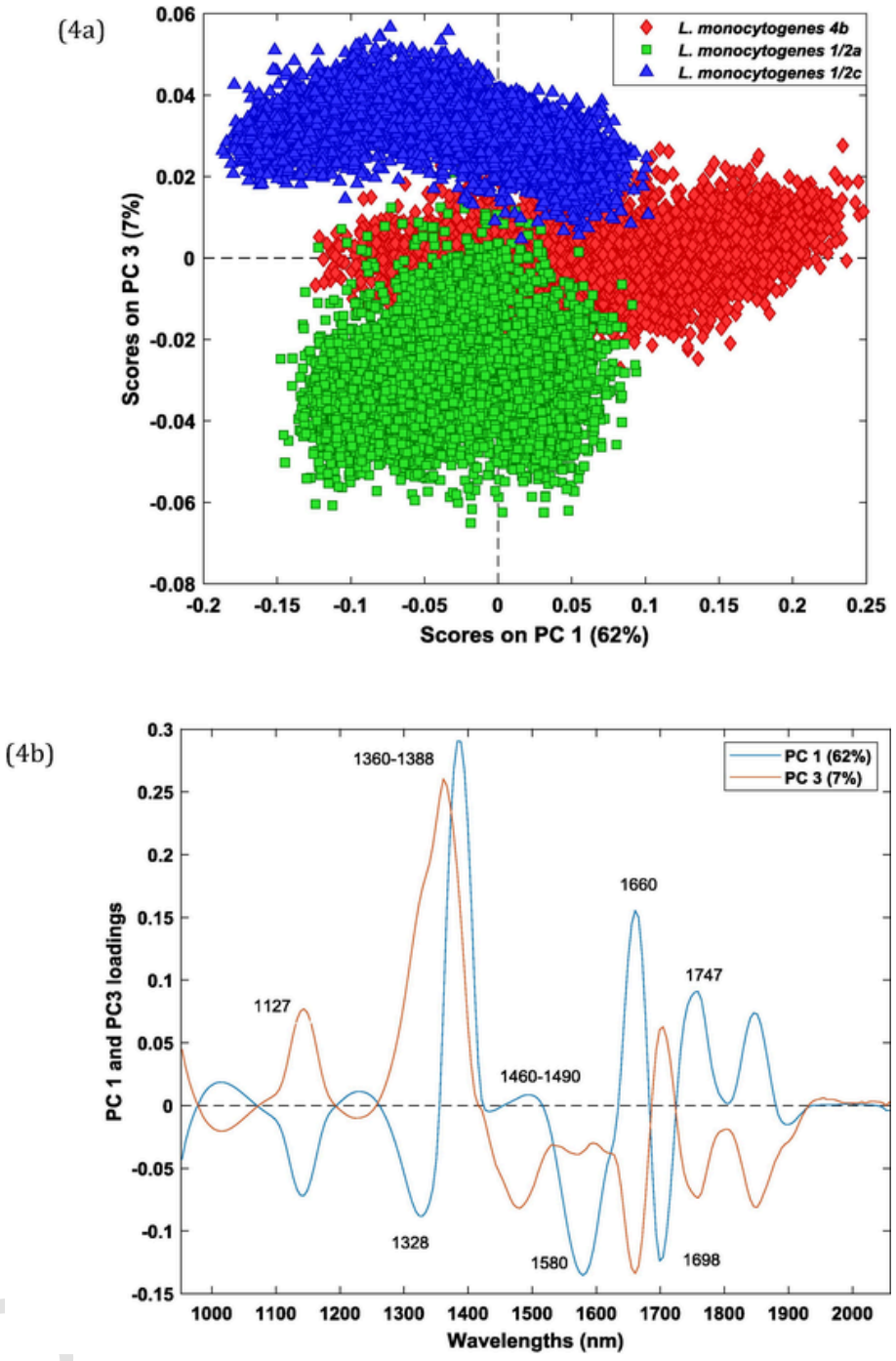

Fig. 4. (a) Principal component analysis score plot of PC1 and PC3 contributing 69% of total variance for *L. monocytogenes* serotype 4b (red), 1/2a (green) and 1/2c (blue) with minimal separation observed. (b) Corresponding PC1 and PC3 loading line plots. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

telluride (HgCdTe) detector, combined with a computer equipped with Breeze software version 2021.1.5 (Prediktera, AB, Umeå, Sweden). Two 150 W halogen lamps (Ushio lighting Inc., Japan) were placed at a 45^0 angle to illuminate the sample. The petri dishes of 100 mm diameter were imaged with the lid opened on a white ceramic tile. The translation stage on which petri dishes were placed moved at a speed of 50 mm/s under the camera equipped with a 30 cm focal length lens. The distance between the samples and the lens was approximately 25 cm. The instrument had a field of view of 95 mm and a spatial resolution of 310 μm. The images were recorded in the 953–2500 nm wavelength range with a spectral resolution of 5.45 nm.

2.3. Hyperspectral image analysis

Hyperspectral images were analysed using Evince v.2.7.0 (Prediktera AB, Umeå, Sweden). Prior to image acquisition, a dark reference (0 % reflectance, camera shutter closed) and a grey reference (50 % reflectance) image was collected by scanning the grey Zenith Allucore diffuse reflectance standard (SphereOptics GmbH, Germany). The 50 %

Table 2
Performance of the PLS-DA models used to distinguish between the three *L. monocytogenes* serotypes using different pre-processing techniques.

		Cross-validation			Prediction		
Preprocessing	Number of LV	Accuracy %	Sensitivity	Specificity	Accuracy %	Sensitivity	Specificity
None	6	68	0.99	0.99	68	0.68	0.84
SNV	6	96	0.96	0.98	85	0.85	0.92
MSC	7	96	0.96	0.98	85	0.85	0.92
1D	6	98	0.98	0.99	74	0.74	0.87
2D	8	91	0.91	0.95	76	0.76	0.88
SNV1D	**5**	**94**	**0.95**	**0.97**	**82**	**0.82**	**0.91**
MSC1D	6	94	0.94	0.97	81	0.81	0.90
SNV2D	9	87	0.87	0.93	81	0.82	0.90
MSC2D	9	88	0.88	0.94	87	0.87	0.93

Table 3
The pixel wise confusion matrix of the PLS-DA model for cross-validation and prediction with respect to the differentiation of *L. monocytogenes* serotypes.

	L. monocytogenes 4b	*L. monocytogenes* 1/2a	*L. monocytogenes* 1/2c	OCCR (%)
Cross-validation (n = 9654 pixels)				94
L. monocytogenes 4b	**3165**	94	111	
L. monocytogenes 1/2a	93	**2974**	13	
L. monocytogenes 1/2c	174	6	**3024**	
Validation (n = 7690 pixels)				82
L. monocytogenes 4b	**1756**	545	46	
L. monocytogenes 1/2a	6	**2195**	0	
L. monocytogenes 1/2c	763	9	**2370**	

Table 4
Performance indices of selected PLS-DA model indicating its ability to predict *Listeria* serotypes.

	Serotype classes		
	L. monocytogenes 4b	*L. monocytogenes* 1/2a	*L. monocytogenes* 1/2c
Cross-validation			
Specificity	0.98	0.98	0.97
Sensitivity	0.92	0.97	0.96
Classification Error (%)	5	2	3
Validation			
Specificity	0.88	0.99	0.85
Sensitivity	0.69	0.80	0.98
Classification Error (%)	18	7	11

grey standard allowed for longer integration times allowing for improved signal to noise ratio. Furthermore, the calculated reflectance values of the samples have been shown to not differ significantly when using the 50 % (grey) instead of the 99 % (white) reflectance standard [29,30]. Following image acquisition, the spectra were converted from reflectance to pseudo absorbance. The Evince software was used to calibrate and correct images according to equation (1).

$$C_{\lambda,n} = -\log_{10}\left[\left(\frac{R_{\lambda,n} - B_{\lambda,n}}{G_{\lambda,n} - B_{\lambda,n}}\right)\right] \quad (1)$$

In the reorganized hypercube, n represents the pixel index variable ($n = 1 \ldots N$). C_λ represents the corrected absorbance image of pixel n at wavelength λ. R_λ represents the sample image of pixel n at wavelength λ. B_λ represents the dark reference image of pixel n at wavelength λ. G_λ represents the grey reference image of pixel n at wavelength λ.

2.4. Data analysis

2.4.1. Region of interest identification and data extraction

Isolation of the region of interests (bacterial colonies) was achieved by removing unwanted pixels such as the background, agar, and shadows from each individual image. This was done using a PCA model, which was calculated on mean centered data with six principal components. The score plots were used to visualize clustering of the different spectral information [31]. The brushing technique was used for removal of unwanted pixels[20,32]. The technique involves eliminating clusters in the score images corresponding to undesired information, followed by recalculating the PCA. The process is repeated iteratively until all undesired information is eliminated, while maintaining the necessary information (Fig. 2). Once the images had been cleaned, spectral data was extracted for analysis in MATLAB R2022b (The MathWorks, MA, USA) and PLS Toolbox version 9.2 (Eigenvector Technologies, USA).

2.4.2. Exploratory and pre-process analysis

Principal component analysis (PCA) was employed for exploratory data analysis and interpretation. PCA is a dimensionality reduction technique which extracts the most valuable information from multidimensional data [33]. In situations where data points or features exhibit redundancy or unclear separation, PCA creates orthogonal components that effectively reduce this overlap, making it easier to identify and understand patterns within the data. After inspecting the spectra, wavelengths between 2100 and 2500 nm were excluded as these were noisy. For the enhancement of spectral signals and reduction of noise, different pre-processing methods were investigated. Standard normal variate (SNV) was applied to correct light scattering effects and baseline offsets. To achieve this, spectral datasets are scaled to have a mean of zero and a standard deviation of one for each wavelength [34]. The treatment was followed by Savitzky-Golay (SG) 1st derivative smoothing filter, 2nd order polynomial, 11-point smoothing, aimed at reducing spectral noise. According to Savitzky and Golay [35], the SG algorithm achieves data smoothing through the minimization of the least squares polynomial approximation, thus improving precision, and making data much easier to analyse and interpret Evaluation of score plots, score images, and loading line plots of pixel data was conducted to investigate potential differences between the bacteria.

2.5. Discriminant analysis

PLS-DA, a supervised linear discriminant technique, was used to build predictive models for the classification of the three *L. monocytogenes* serotypes. The algorithm works by obtaining latent variables that are linear combinations of the original predictor variables [36]. Choos-

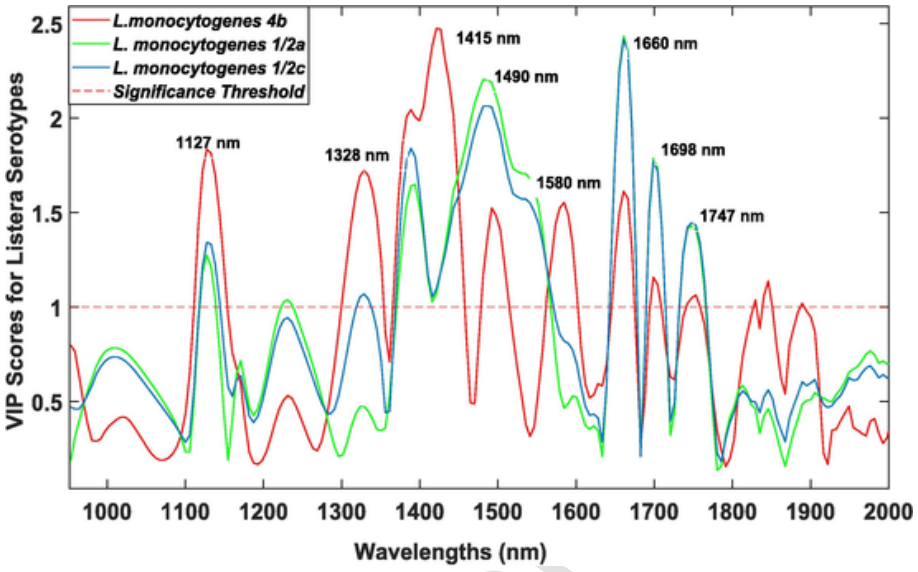

Fig. 5. VIP scores for *L. monocytogenes* serotypes showing important peaks at 1127 nm, 1328 nm, 1415 nm, 1490 nm, 1580 nm, 1660 nm, 1698 nm, and 1747 nm.

ing these latent variables maximizes the covariance between predictor variables and class labels. Based on the predictor variables, new observations can then be categorized into predefined groups [37]. In this study, the pixel-wise analysis method was explored. Each pixel was regarded as a sample, hence contributing a single spectrum. Due to the large amount of pixels involved in pixel-wise analysis, using an external set for validation is highly recommended and is easily obtainable [38]. Hence, the petri dishes obtained from the first experiment (Rep 1) were chosen as the calibration set (train set) and those from the second experiment (Rep 2) as the validation set (test set). The calibration set used to train and perform cross-validation consisted of spectra from 9 654 pixels, whilst the test set used for validation had spectra from 7 690 pixels. Table 1 shows the composition of the data set.

Due to the complexity of the data and the risk of model overfitting, cross-validation (CV) was necessary [36]. The venetian blind method was applied with 10 data splits. In this method, the data set is first divided into equal sized folds (k) and at each iteration one is used as the test set while the remaining k-1 folds are used as the training set [39]. The method is suitable for large data sets arranged in a random order and for assessing predictability within batches. The number of latent variables was chosen by considering both the calibration (CAL) and CV classification average error. The latent variable with the minimum error rate supports the model's predictive performance or accuracy and is less prone to overfitting.

There are several parameters that can be used to assess a classification model's performance, including but not limited to efficiency, Matthew's correlation coefficient, accuracy, and classification error [36,40]. However, in this study, classification error, specificity, sensitivity, and classification accuracy were used. A model's classification accuracy illustrates the model's overall performance. This is calculated by dividing the number of correctly classified objects by the total number of predictions (Equation 2), [41]. However, it is worth noting that classification accuracy should not be regarded as the sole determinant of performance. It is conceivable to achieve a reasonably high classification accuracy and yet fail to correctly identify the target class. Hence, models often evaluate other parameters such as sensitivity and specificity. The sensitivity of a model measures its ability to accurately assign objects to their respective classes. Specificity on the other hand indicates how well a model detects the number of samples that are correctly predicted to the negative class (true negatives). Sensitivity and specificity values above 0.8 are generally considered good however, the optimal values may vary depending on the specific requirements of the test [42].

$$\text{Overall correct classification rate (OCCRs)} = \frac{Correctnumberofspectra/pixels}{Totalnumberofspectra/pixels} \times 100\%$$

(Equation 2)

2.6. Variable selection

The process of model building involves assessing and selecting the most relevant variables, which play a crucial role on the model's ability to better capture patterns and relationships within the data. Using variable selection methods contribute to enhancement of the model's performance, by effectively reducing noise caused by irrelevant features [43]. Hence, variable importance in projection (VIP) scores were evaluated to determine the wavebands most important for classifying the *L. monocytogenes* serotypes. In a PLS-DA model, VIP scores are computed by assessing the contribution of each predictor variable to the separation of the classes in the model [44]. To achieve this, one must consider the correlation between predictor variables and compare the amount of variance explained by the variable in question to the overall amount of variance in the model. In general, variables with VIP score values greater than 1 are considered important for the model's performance [45,46]. Equation (3) illustrates the mathematical computation.

$$VIP_a = \sqrt{c\sum_{k=1}^{v} w_{ak}^2 \left(\frac{\sum_{a=1}^{n} t_{sa}^2}{SSY}\right)} \quad (3)$$

where:
VIP_a = VIP score for variable *a*,
c = number of components of the pls model,
v = the number of response variables,
w_{ak} = weight of variable a in component k,
t_{sa} = the score of sample *s* on component a,
SSY = total sum of squares of the response variables.

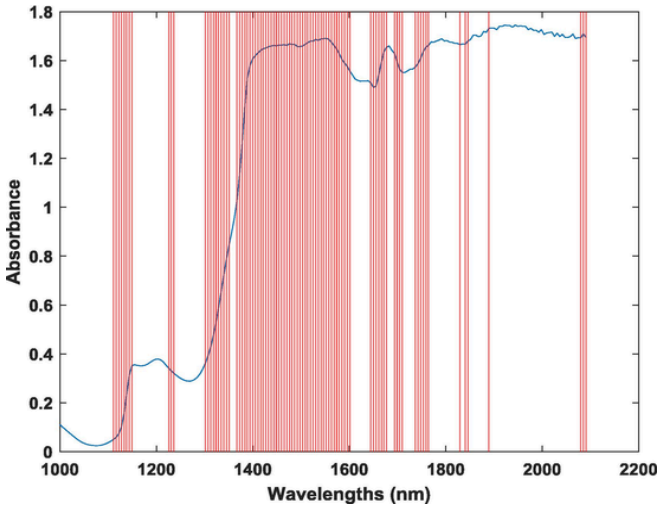

Fig. 6. Graphical representation of the variables selected using variable importance in projection feature selection.

3. Results and discussion

3.1. NIR spectra

The raw absorbance spectra of the three *L. monocytogenes* serotypes grown on BHI agar, are shown in Fig. 3. The spectral profiles displayed a relative similarity (in shape), but with notable differences in the wavelength range 1100–1800 nm. Considering Fig. 3a, in the wavelength range 1395 to 1410 nm, there seems to be high absorbance values which could be a resultant of O–H stretch overtones or C–H combination stretching likely from water or sugars, respectively [47]. There were also considerable peaks at 1154 and 1682 nm, however information from the raw spectra was insufficient to draw any meaningful conclusions, hence the need to pre-process spectral data. The inadequacy to capture interpretable, detailed spectral information is often attributed to the negative impact of both light scattering and spectral noise on NIR radiation [48]. To minimise these effects, spectral data was subjected to SNV and S.G smoothing treatment 2nd order polynomial, 1st derivative 11-point window both independently and in combination. When comparing the pre-processed spectra of SNV and SG in combination versus the spectra with only the SG 1st derivative, no noticeable differences were observed. Hence information derived from the first derivative will be discussed. Overall, Fig. 3b shows more detail highlighting the effectiveness of pre-processing on spectral data.

Fig. 3b shows wavebands 1018 nm, 1127 nm, and 1370 nm resultant from N-H, C–H stretch second overtone and a combination of C–H deformation and C–H-stretching respectively, all associated with amino sugars [49,50]. A prominent peak at 1370 nm is due to a combination C–H stretch. The wave band 1530 nm and 1580 nm are due to O–H 1st overtone intermolecular bond known to be associated with carbohydrates and simple sugars [51]. Wavebands 1660 nm, 1698 nm and 1747 nm are resultant of C–H stretching first overtone mostly associated with hydrocarbons which could be monosaccharides or fatty acids [52]. Upon further investigation, the spectra of *L. monocytogenes* serotypes 1/2a and 1/2c appear similar, indicating a strong relationship between the two serotypes. The spectra of serotype 4b is slightly different from the other two suggestive that there are differences between the two lineage groups. The differentiation of *L. monocytogenes* into lineages is based on the presence of unique genomic segments that are specific to each lineage [53]. Both *L. monocytogenes* serotypes 1/2a and 1/2c belong to lineage II and have similar O-antigens which are potentially different from serotype 4b, explaining the patterns observed [54,55]. The minor structural variations of peptidoglycan wall-teichoic acids (WTAs) and specific glycosylation patterns define O-antigens responsible for serotyping [55]. According to Kamisango *et al.* [56], the genetic differences in *L. monocytogenes* lineages can affect the composition and structure of the cell wall. Previous studies have shown that *L. monocytogenes* serotype 4b teichoic acid structures have more galactose when compared with serotypes 1/2a and 1/2c [57,58]. Hence suggesting that the dissimilarities between the three serotypes primarily arose from variations in carbohydrates within the bacterial cell wall. However, given that all three serotypes stem from the same bacterial species, a similar shape of the spectra observed was also expected.

3.2. PCA results

After evaluating different pre-processing algorithms, SNV and Savitzky Golay 1st derivative, 2nd order polynomial 11-point smoothing were applied. The resulting 2D score plot (Fig. 4a) with six principal components, showed 62 % variation along PC1 and 7 % variation along PC3. There are three clusters, with a clear separation between *L. monocytogenes* 1/2a (green) and *L. monocytogenes* 1/2c (blue) along PC3. Additionally, there is partial separation along PC1 between lineage I (serotype 4b) and lineage II (serotype 1/2a and serotype 1/2c). The lack of distinct boundaries between the three serotypes along PC1 and PC3 implies a degree of similarity among them. This is consistent with the similar spectra shape of the three serotypes observed (Fig. 3a). However, the partial separation of all three serotypes along PC1 might be resultant of variations within the structure of the cell wall [59].

The corresponding loadings for PC1 and PC3 are illustrated in Fig. 4b. The two loadings were examined to identify chemical information contributing to the observed clustering. In the loading plots of PC1 and PC3, a notable positive feature appears at 1360–1388 nm, indicating the potential presence of hydrocarbons (C–H combination stretches). This observation may be associated with the sugar content from both the bacteria and agar. This variance contribution could explain the overlap of the three serotypes as observed in the PCA score plots.

PC1 loading showed wavebands 1127 nm, 1328 nm, 1460–1490 nm, 1580 nm, 1660, 1698 nm, and 1747 nm as the major contributors to the patterns observed in the score plots. The positively loaded wavebands exhibited a stronger association with *L. monocytogenes* 4b pixels in comparison to *L. monocytogenes* 1/2a and 1/2c pixels. Bands 1660, and 1747 nm can be ascribed to C–H stretching vibrations and C–H bond first overtone respectively, arising from hydrocarbon chains of sugars. Wavebands 1460–1490 nm cover a wide range N–H 1st overtone stretch and intramolecular O–H stretches. Variance contribution could be from amino sugars and their interaction with water molecules. The negatively loaded wavebands of PC1 include 1328 nm, 1580 nm, and 1698 nm. These demonstrated a more pronounced correlation with *L. monocytogenes* 1/2a and 1/2c pixels compared to *L. monocytogenes* 4b pixels. Waveband 1580 nm is strongly associated with the vibrations of intermolecular O–H bonds [60]. This stretching vibration is indicative of the presence of carbohydrates or differences in water content in the cell walls of all three *L. monocytogenes* serotypes, attributing to the overlapping pixels observed in the PCA score plots.

Variance attributed to PC3, was associated with wavebands: 1127 nm, 1490 nm, 1660 nm, and 1698 nm. The loading plot suggested that the main variation source for *L. monocytogenes* 1/2a was in the wavelength range 1400–1800 nm. According to Barbin *et al.* [61], bands at 1660 nm can be attributed to the presence of a C–H bond first overtone from the CH_3 group from the monosaccharides (sugars) in the cell wall. Variance contribution from positively loaded wavebands which include 1127 nm, 1698 nm are closely associated with a symmetrical N–H stretch (second overtone) and $CONH_2$ respectively. These potentially stem from amino groups within the cell wall [19,60]. Furthermore, waveband 1360 nm can be attributed to the CH_3 structure re-

Table 5
Reduced wavelength PLS-DA model performance measures for the *Listeria* serotypes.

	Serotype classes		
	L. monocytogenes 4b	*L. monocytogenes* 1/2a	*L. monocytogenes* 1/2c
Cross −validation			
Specificity	0.97	0.98	0.98
Sensitivity	0.93	0.96	0.98
Classification Error (%)	4	3	2
Validation			
Specificity	0.92	0.99	0.83
Sensitivity	0.65	0.85	0.98
Classification Error (%)	17	5	12

sulting from both C–H stretch and C–H deformation bonds [51]. However, it is worth noting that since there still is minimal overlapping in the score plots along PC3, all the resultant positive loading peaks might potentially be from all serotypes. Additionally, the observed contributions could be resultant of other R-OH related bonds such as carbohydrates. Cell walls of Gram-positive bacteria contain chains of teichoic acid that extend from the plasma membrane to the surface of the cell wall through the peptidoglycan layer [62]. Hence, the peaks associated with C–H and O–H stretches observed in the loadings can be explained by the presence of sugars like glycerol or ribitol which are the main components of teichoic acids. Furthermore, since the peptidoglycan layer has covalently bonded surface proteins, this explains the peak at 1127 nm and 1490 nm which is suggestive of differences in protein type between *L. monocytogenes* serotype 1/2a and 1/2c. Past research on other Gram-positive bacteria like *S. aureus* has shown that differences in protein content among strains is possible [63,64]. However, there is limited literature concerning *L. monocytogenes* serotypes, leaving a gap worth further investigation.

The varying loadings at different wavelengths are indicative that specific spectral regions can provide informative on distinguishing between different classes. However, it is also important to note that the substrate (BHI agar) on which the bacteria grew could have contributed to the overall spectral characteristics of the bacteria [65]. BHI agar constitutes of glucose and peptone which could also have contributed to some of the peaks observed. Nevertheless, the spectral characteristics of the bacterial colonies including those influenced by the agar have been proven to be successfully used in predictive model development [27]. Hence, to further validate the findings, PLS-DA was employed.

3.3. Classification results

3.3.1. Full wavelength PLS-DA model

A variety of pre-processing techniques were examined for optimal PLS-DA model development. Ultimately, SNV and Savitzky Golay 1st derivative 2nd polynomial 11-point smoothing yielded the best classification results with the least number of latent variables (Table 2). Furthermore, the Q residuals were also considered, with the pre-processing technique with the lowest Q residual values being accepted. High Q residual values indicate a large discrepancy between actual and predicted labels, resulting in a poor performing model [66]. To identify the optimal number of latent variables to use in the model, the graphical plot of classification error vs latent variable number was considered. The optimal number of latent variables was determined by selecting the point where the classification error rate is minimized or stabilised [67]. Hence, for optimum classification accuracy, 5 latent variables were adopted. This choice resulted in a minimum average classification error rate of 0.07, as increasing the number of latent variables beyond 6 did not reduce the error rate.

Table 3 presents a confusion matrix that illustrates pixel-wise classification of the microbial classes. The PLS-DA model gave satisfactory discrimination results, with overall correct classification rates of 94 % and 82 % for cross-validation and external validation, respectively. The performance indices of PLS-DA models, which indicate the model's ability to accurately predict and classify unknown *L. monocytogenes* serotypes, are presented in Table 4. Cross-validation results revealed high specificity values of 0.98, 0.98 and 0.97 for *L. monocytogenes* serotypes 4b, 1/2a, and 1/2c, respectively (Table 4). The model's sensitivity was also high, with 0.97 and 0.96 for *L. monocytogenes* serotypes 1/2a, and 1/2c, respectively. However, it is also important to note that for *L. monocytogenes* 4b, the sensitivity value was 0.92 which was indicative that the model has difficulty identifying true 4b serotypes. The classification error for the serotype 4b is also quite high at 5 %, as compared with 2 % and 3 % for serotype 1/2a and 1/2c, respectively.

Validation results also demonstrated a similar trend with *L. monocytogenes* 1/2a and *L. monocytogenes* 1/2c having high sensitivity values of 0.80 and 0.98 respectively. However, serotype 4b had a lower sensitivity value of 0.69. The results suggest that the model is able to correctly identify the bacteria of choice (true positives, TP) better for lineage II serotypes compared to lineage I (*L. monocytogenes* 4b). In addition, the results also showed a high specificity value of 0.99 for *L. monocytogenes* 1/2a suggesting that the model can accurately predict the true negatives (TN). However, the specificity values of *L. monocytogenes* 1/2c and *L. monocytogenes* 4b were slightly lower, with 0.85 and 0.88 respectively. This indicates that the model has difficulties identifying true negatives and has a high number of false positives. Whether a sensitivity or specificity value is acceptable or not is dependent on the context of the analysis and the goals of the study. However, generally, sensitivity and specificity values greater than 0.80 are considered good and those below 0.70 considered poor [45,68]. In the food industry, a high level of sensitivity is desired for pathogenic bacteria such as *L. monocytogenes* detection. Therefore, a sensitivity value of 0.69 is not considered good enough for adoption. Despite the model achieving an overall classification accuracy of 82 % in the validation results, it is worth noting that all serotypes exhibited classification errors exceeding 5 %. This suggests that the model's performance in accurately classifying or predicting *L. monocytogenes* in new datasets, particularly in the case of *L. monocytogenes* 4b, was suboptimal. In general, sensitivity and specificity values of above 0.90 and 0.95 respectively are acceptable for *L. monocytogenes* detection methods [69].

3.4. VIP scores

VIP scores were employed to identify the wavebands that had the highest influence on discrimination between the serotypes (Fig. 5). A visual inspection showed that the wavelength region 1100–1800 nm was important in modeling the three groups of bacteria. In the context of the model, wavebands that possessed VIP scores greater than one were considered significant [70]. The wavebands identified and selected were 1127 nm, 1328 nm, 1415 nm, 1490 nm, 1580 nm, 1660 nm, 1698 nm, and 1747 nm. The wavebands 1490 nm, 1580 nm, 1660 nm, 1698 nm, and 1747 nm exhibited greater importance in distinguishing lineage II serotypes (1/2a and 1/2c) compared to lineage I (4b) serotypes. On the contrary, wavelengths 1127 nm, 1328 nm and 1415 nm demonstrated to be more important for *L. monocytogenes* 4b. It is noteworthy that the wavelengths selected were consistent with the insights gained from the PC loadings, suggesting that separation between the serotypes was mainly based on variation in protein (1490 nm) and carbohydrates (1580 and 1698 nm) content.

However, to get a more comprehensive assessment and to determine if reduction in variables would optimise the model's performance, an in-house algorithm was applied to automatically select all variables with a VIP score above 1. This selection yielded 89 variables (Fig. 6)

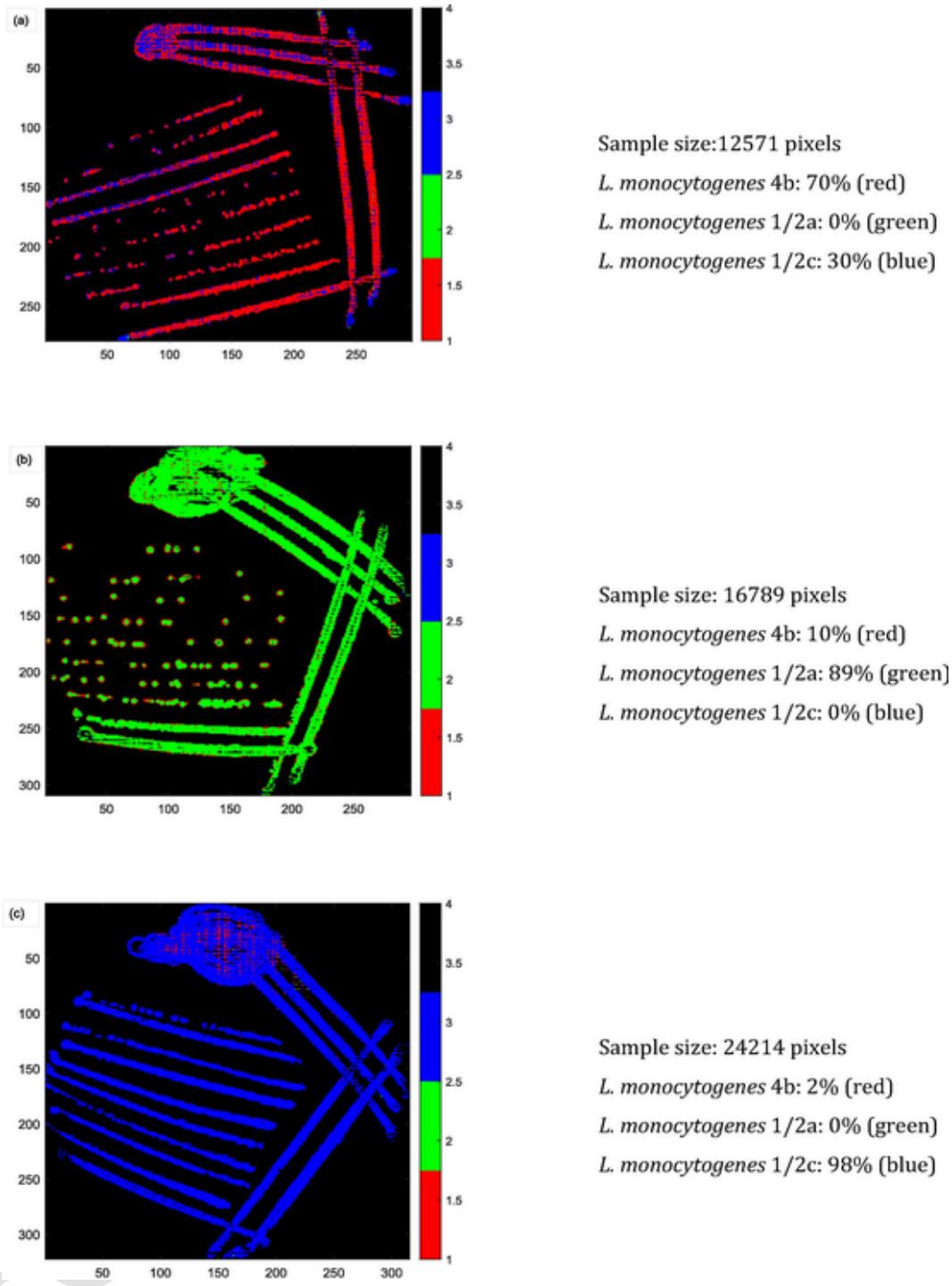

Fig. 7. Prediction maps generated using the full wavelength partial least discriminant analysis model on hyperspectral images of: (a) *L. monocytogenes* 4b (red), (b) *L. monocytogenes* 1/2a (green) and (c) *L. monocytogenes* 1/2c (blue) and the corresponding prediction results. Pixels classified as agar/background are represented in black. The colour bar indicates the predicted class assignment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

which were then used to develop a reduced wavelength PLS-DA model. Table 5 shows the performance measures for the *Listeria* serotypes.

The model's overall classification accuracy was 94 % and 82 % for the cross validation and validation sets. Cross-validation results revealed high specificity values of 0.97, 0.98 and 0.98 for *L. monocytogenes* serotypes 4b, 1/2a, and 1/2c, respectively. This was relatively similar to the full wavelength results. The model's sensitivity was also high, with 0.93, 0.96 and 0.98 for *L. monocytogenes* serotypes 4b, 1/2a, and 1/2c, respectively. The classification errors slightly improved for serotype 4b and 1/2c with values of 4 %, and 2 %, respectively. The classification error for serotype 1/2a increased to 3 %.

The validation results indicated a decline in the model's performance when an independent test set was introduced. The sensitivity value for serotype 4b was 0.65 while for serotype 1/2a, and 1/2c it was

0.85 and 0.98 respectively. The specificity values were 0.92, 0.99 and 0.83 for serotype 4b, 1/2a and 1/2c, respectively. The low specificity values for serotype 1/2c are indicative that the model has difficulties identifying true negatives and has a high number of false positives. Moreover, the classification errors for serotype 4b and 1/2c were high at 17 % and 12 % respectively.

The performance of the reduced wavelength PLS-DA model did not show a substantial improvement over the full wavelength model. When comparing the results between the full wavelength range and the limited variable set, both scenarios yielded similar outcomes. The model achieved an overall classification accuracy of 94 % and 82 % for cross-validation data and validation data. This outcome suggests that the selected subset is still capturing the essential information necessary for effective discrimination.

3.5. Prediction maps for L. Monocytogenes serotypes

One advantage of hyperspectral imaging over conventional spectroscopy is its ability to utilize spatial data, which allows for the generation of distribution maps to visualize spatial patterns. Prediction maps are spatial representations generated through statistical or mathematical models, illustrating the predicted values of a variable investigated [71]. By leveraging both the spatial and spectral information captured across numerous wavelengths, these maps offer a comprehensive visualization of model classification or predictions.

Fig. 7 shows prediction maps of the different *L. monocytogenes* serotypes using the full wavelength PLS-DA model on validation data hypercubes. These maps depict the predicted distributions across the entire agar plates alongside their corresponding prediction results. A linear colour scale from red to black on the right aids in interpretation. In these maps, class 1 (*L. monocytogenes* 4b) is depicted in red (Fig. 7a); class 2 (*L. monocytogenes* 1/2a) appears in green (Fig. 7b), and class 3 (*L. monocytogenes* 1/2c) is shown in blue (Fig. 7c). Class 4 (agar/ background) is represented in black.

Fig. 7b and 7c, depicting serotypes 1/2a and 1/2c in green and blue, respectively, showcase accurate prediction with distinct colour separation. This clarity suggests good model performance in identifying these serotypes across the plates. However, the representation of serotype 4b (red) appears less defined, with observed instances of misclassification as serotype 1/2c pixels (30 %). This discrepancy aligns with the high misclassification rates and low sensitivity values documented in Table 4, suggesting weaker model performance in accurately identifying serotype 4b.

4. Conclusion

In this study, NIR hyperspectral imaging was investigated for the detection and classification of *L. monocytogenes* serotypes 1/2a, 1/2c and 4b. Pixel-wise analysis was adopted based on the ability to provide and retain more spectral detail. Spectral results demonstrated minimum variation, with the spectra obtained having a similar shape. This was due to the related biological and serological properties of the *L. monocytogenes* serotypes. Loading plots revealed the contributions of individual variables to specific components, while the score plots displayed partial separation among serotypes. Notably, variables associated with amino acids (1490 nm) and sugars (1580 nm) play a role in this separation. Further investigation with PLS-DA showed feasible classification of the serotypes. The full wavelength model had overall sensitivity, specificity, and classification accuracies above 0.87, 0.85, 80 % respectively. While the reduced wavelength model had overall sensitivity, specificity, and classification accuracies of approximately 0.83, 0.90, 88 % respectively. It is important to note that despite the relatively favourable sensitivity and specificity values, these might not meet the acceptable threshold for pathogenic bacteria detection as this typically demands values above 90 %. However, considering the study's aim was to evaluate the potential use of NIR-HSI for bacteria detection rather than replacing existing methods, these results are acceptable. Based on this perspective, it is recommended that this method be used as an early detection system. The models developed could be integrated with traditional microbial assessment techniques to further confirm and give a comprehensive assessment of bacterial identification and quantification. In addition, further research should also include different growth media and the application of other modelling chemometric tools to ensure a comprehensive assessment of serotype classification and a more robust accurate model.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

CRediT authorship contribution statement

Rumbidzai T. Matenda: Writing – original draft, Project administration, Methodology, Formal analysis, Conceptualization. **Diane Rip:** Writing – review & editing, Supervision, Methodology, Conceptualization. **J.A. Fernández Pierna:** Writing – review & editing, Supervision. **Vincent Baeten:** Writing – review & editing, Supervision. **Paul J. Williams:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [Paul James Williams reports financial support was provided by National Research Foundation of South Africa. Rumbidzai T Matenda reports financial support and travel were provided by South Africa Department of Science and Innovation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.].

Data availability

Data will be made available on request.

Acknowledgments

This work is based on the research supported wholly / in part by the National Research Foundation of South Africa (Grant Number: 137998). The authors would also like to acknowledge Nicaise Kayoka Mukendi from the Walloon Agricultural research Centre (CRA-W) in Gembloux, Belgium, for his assistance in data preparation and analysis. The authors would also like to acknowledge the Department of Science and Innovation (South Africa) in collaboration with the SensorFint Cost Action 19145 for financial and training support. In addition, Stellenbosch University Postgraduate Scholarship Programme for their financial support to the student Rumbidzai T. Matenda.

References

[1] World Health Organisation, “WHO estimates of the global burden of foodborne dieseases:foodborne diseases burden epidemiology reference group 2007-2015.” pp. 1–1, 2015. doi: Doi: 10.1007/978-3-642-27769-6_3884-1.
[2] K. Z. Bisholo, S. Ghuman, and F. Haffejee, “Food-borne disease prevalence in rural villages in the Eastern Cape, South Africa,” *African J. Prim. Heal. Care Fam. Med.*, vol. 10, no. 1, p. 5, 2018, [Online]. Available: https://phcfm.org/index.php/phcfm/article/view/1796/2825.
[3] E. Abebe, G. Gugsa, M. Ahmed, Review on Major Food-Borne Zoonotic Bacterial Pathogens, J. Trop. Med. 2020 (2020), https://doi.org/10.1155/2020/4674235.
[4] J. Thomas, et al., Outbreak of Listeriosis in South Africa Associated with

Processed Meat, N. Engl. J. Med. 382 (7) (2020) 632–643, https://doi.org/10.1056/nejmoa1907462.
[5] J. Lundén, R. Tolvanen, H. Korkeala, Human listeriosis outbreaks linked to dairy products in Europe, J. Dairy Sci. 87 (SUPPL. 1) (2004) 6–12, https://doi.org/10.3168/jds.S0022-0302(04)70056-9.
[6] J. A. Gray, P. S. Chandry, M. Kaur, C. Kocharunchitt, J. P. Bowman, and E. M. Fox, "Novel biocontrol methods for Listeria monocytogenes biofilms in food production facilities," *Front. Microbiol.*, vol. 9, no. APR, pp. 1–12, 2018, doi: 10.3389/fmicb.2018.00605.
[7] R.H. Orsi, M. Wiedmann, Characteristics and distribution of Listeria spp., including Listeria species newly described since 2009, Applied Microbiology and Biotechnology 100 (12) (2016) 5273–5287, https://doi.org/10.1007/s00253-016-7552-2.
[8] P. Wang, et al., Characterization of Listeria Monocytogenes Isolates in Import Food Products of China from 8 Provinces Between 2005 and 2007, J. Food Sci. 77 (4) (2012) pp, https://doi.org/10.1111/j.1750-3841.2011.02596.x.
[9] R.H. Orsi, H.C. de. Bakker, M. Wiedmann, Listeria monocytogenes lineages: Genomics, evolution, ecology, and phenotypic characteristics, Int. J. Med. Microbiol. 301 (2) (2011) 79–96, https://doi.org/10.1016/j.ijmm.2010.05.002.
[10] G.T. Jeffers, J.L. Bruce, P.L. McDonough, J. Scarlett, K.J. Boor, M. Wiedmann, Comparative genetic characterization of Listeria monocytogenes isolates from human and animal listeriosis cases, Microbiology 147 (5) (2001) 1095–1104, https://doi.org/10.1099/00221287-147-5-1095.
[11] S. V. Poimenidou, M. Dalmasso, K. Papadimitriou, E. M. Fox, P. N. Skandamis, and K. Jordan, "Virulence gene sequencing highlights similarities and differences in sequences in Listeria monocytogenes serotype 1/2a and 4b strains of clinical and food origin from 3 different geographic locations," *Front. Microbiol.*, vol. 9, no. JUN, 2018, doi: 10.3389/fmicb.2018.01103.
[12] V. Braga, et al., Prevalence and serotype distribution of Listeria monocytogenes isolated from foods in Montevideo-Uruguay, Brazilian J. Microbiol. 48 (4) (2017) 689–694, https://doi.org/10.1016/j.bjm.2017.01.010.
[13] H. Momtaz, S. Yadollahi, Molecular characterization of Listeria monocytogenes isolated from fresh seafood samples in Iran, Diagn. Pathol. 8 (1) (2013) 1–6, https://doi.org/10.1186/1746-1596-8-149.
[14] V. Ramaswamy *et al.*, "Listeria — review of epidemiology and pathogenesis," pp. 4–13, 2007.
[15] J.W.F. Law, N.S.A. Mutalib, K.G. Chan, L.H. Lee, Rapid metho ds for the detection of foodborne bacterial pathogens: Principles, applications, advantages and limitations, Front. Microbiol. 5, DEC (2014) 1–19, https://doi.org/10.3389/fmicb.2014.00770.
[16] X. Zhao, C.W. Lin, J. Wang, D.H. Oh, Advances in rapid detection methods for foodborne pathogens, J. Microbiol. Biotechnol. 24 (3) (2014) 297–312, https://doi.org/10.4014/jmb.1310.10013.
[17] M.K. Grewal, P. Jaiswal, S.N. Jha, Detection of poultry meat specific bacteria using FTIR spectroscopy and chemometrics, J. Food Sci. Technol. 52 (6) (2015) 3859–3869, https://doi.org/10.1007/s13197-014-1457-9.
[18] A. Aït-Kaddour, T. Boubellouta, I. Chevallier, Development of a portable spectrofluorimeter for measuring the microbial spoilage of minced beef, Meat Sci. 88 (4) (2011) 675–681, https://doi.org/10.1016/j.meatsci.2011.02.027.
[19] Y.Z. Feng, G. Downey, D.W. Sun, D. Walsh, J.L. Xu, Towards improvement in classification of Escherichia coli, Listeria innocua and their strains in isolated systems based on chemometric analysis of visible and near-infrared spectroscopic data, J. Food Eng. 149 (2015) 87–96, https://doi.org/10.1016/j.jfoodeng.2014.09.016.
[20] T.L. Kammies, M. Manley, P.A. Gouws, P.J. Williams, Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis, Appl. Microbiol. Biotechnol. 100 (21) (2016) 9305–9320, https://doi.org/10.1007/s00253-016-7801-4.
[21] K.X. Mu, Y.Z. Feng, W. Chen, W. Yu, Near infrared spectroscopy for classification of bacterial pathogen strains based on spectral transforms and machine learning, Chemom. Intell. Lab. Syst. 179 (2018) 46–53, https://doi.org/10.1016/j.chemolab.2018.06.003.
[22] D. Li, F. Zhang, J. Yu, X. Chen, B. Liu, and X. Meng, "A rapid and non-destructive detection of Escherichia coli on the surface of fresh-cut potato slices and application using hyperspectral imaging," *Postharvest Biol. Technol.*, vol. 171, no. September 2020, p. 111352, 2021, doi: 10.1016/j.postharvbio.2020.111352.
[23] A. Soni, Y. Dixit, M.M. Reis, G. Brightwell, Hyperspectral imaging and machine learning in food microbiology: Developments and challenges in detection of bacterial, fungal, and viral contaminants, Compr. Rev. Food Sci. Food Saf. 21 (4) (2022) 3717–3745, https://doi.org/10.1111/1541-4337.12983.
[24] Y.Z. Feng, W. Yu, W. Chen, K.K. Peng, G.F. Jia, Invasive weed optimization for optimizing one-agar-for-all classification of bacterial colonies based on hyperspectral imaging, Sensors Actuators, B Chem. 269 (2018) 264–270, https://doi.org/10.1016/j.snb.2018.05.008.
[25] G. Foca, et al., The potential of spectral and hyperspectral-imaging techniques for bacterial detection in food: A case study on lactic acid bacteria, Talanta 153 (2016) 111–119, https://doi.org/10.1016/j.talanta.2016.02.059.
[26] S.C. Yoon, et al., Hyperspectral imaging for differentiating colonies of non-O157 Shiga-toxin producing Escherichia coli (STEC) serogroups on spread plates of pure cultures, J. near Infrared Spectrosc. 21 (2) (2013) 81–95, https://doi.org/10.1255/jnirs.1043.
[27] P. Gu *et al.*, "Unified classification of bacterial colonies on different agar media based on hyperspectral imaging and machine learning," *Molecules*, vol. 25, no. 8, Apr. 2020, doi: 10.3390/molecules25081797.
[28] C. H. Feng, Y. Makino, S. Oshita, and J. F. García Martín, "Hyperspectral imaging and multispectral imaging as the novel techniques for detecting defects in raw and processed meat products: Current state-of-the-art research advances," *Food Control*, vol. 84, no. September 2017, pp. 165–176, 2018, doi: 10.1016/j.foodcont.2017.07.013.
[29] K. Edwards, L.C. Hoffman, M. Manley, P.J. Williams, Raw Beef Patty Analysis Using Near-Infrared Hyperspectral Imaging: Identification of Four Patty Categories, Sensors 23 (2) (2023) pp, https://doi.org/10.3390/s23020697.
[30] K. Sendin, M. Manley, F. Marini, P.J. Williams, Hierarchical classification pathway for white maize, defect and foreign material classification using spectral imaging, Microchem. J. 162 (Mar. 2021) 105824, https://doi.org/10.1016/J.MICROC.2020.105824.
[31] K. Esbensen, P. Geladi, Strategy of multivariate image analysis (MIA), Chemom. Intell. Lab. Syst. 7 (1–2) (1989) 67–86, https://doi.org/10.1016/0169-7439(89)80112-1.
[32] E. Bonah, X. Huang, R. Yi, J. H. Aheto, and S. Yu, "Vis-NIR hyperspectral imaging for the classification of bacterial foodborne pathogens based on pixel-wise analysis and a novel CARS-PSO-SVM model," *Infrared Phys. Technol.*, vol. 105, no. December 2019, p. 103220, 2020, doi: 10.1016/j.infrared.2020.103220.
[33] H. Jiang, W. Yuan, Y. Ru, Q. Chen, J. Wang, and H. Zhou, "Feasibility of identifying the authenticity of fresh and cooked mutton kebabs using visible and near-infrared hyperspectral imaging," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 282, no. July, 2022, doi: 10.1016/j.saa.2022.121689.
[34] M.R. Maleki, A.M. Mouazen, H. Ramon, J. De Baerdemaeker, Multiplicative Scatter Correction during On-line Measurement with Near Infrared Spectroscopy, Biosyst. Eng. 96 (3) (2007) 427–433, https://doi.org/10.1016/j.biosystemseng.2006.11.014.
[35] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation," *Anal. Chem*, vol. 36, no. 8, pp. 1627–1639, 1964, [Online]. Available: Doi: 10.1021/ac60214a047.
[36] D. Ruiz-Perez, H. Guan, P. Madhivanan, K. Mathee, G. Narasimhan, So you think you can PLS-DA? BMC Bioinformatics 21 (Suppl 1) (2020) 1–10, https://doi.org/10.1186/s12859-019-3310-7.
[37] R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: Taking the magic away, J. Chemom. 28 (4) (2014) 213–225, https://doi.org/10.1002/cem.2609.
[38] J.M. Amigo, H. Babamoradi, S. Elcoroaristizabal, Hyperspectral image analysis. A tutorial, Anal. Chim. Acta 896 (2015) 34–51, https://doi.org/10.1016/j.aca.2015.09.030.
[39] M. Kuhn and K. Johnson, *Applied Predictive Modeling with Applications in R*, vol. 26. 2013. [Online]. Available: http://appliedpredictivemodeling.com/s/Applied_Predictive_Modeling_in_R.pdf.
[40] J.A. Westerhuis, et al., Assessment of PLSDA cross validation, Metabolomics 4 (1) (2008) 81–89, https://doi.org/10.1007/s11306-007-0099-6.
[41] M. Sokolova, N. Japkowicz, S. Szpakowicz, "Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation", *AAAI Work.* -, Tech. Rep. WS-06-06 (2006) 24–29, https://doi.org/10.1007/11941439_114.
[42] D. Yang, D. He, A. Lu, D. Ren, J. Wang, Detection of the Freshness State of Cooked Beef During Storage Using Hyperspectral Imaging, Appl. Spectrosc. 71 (10) (2017) 2286–2301, https://doi.org/10.1177/0003702817718807.
[43] M. Cocchi, A. Biancolillo, F. Marini, Chemometric Methods for Classification and Feature Selection vol. 82 (2018), https://doi.org/10.1016/bs.coac.2018.08.006.
[44] I.G. Chong, C.H. Jun, Performance of some variable selection methods when multicollinearity is present, Chemom. Intell. Lab. Syst. 78 (1) (2005) 103–112, https://doi.org/10.1016/j.chemolab.2004.12.011.
[45] F.B. De Santana, L.C. Gontijo, H. Mitsutake, S.J. Mazivila, L.M. De Souza, W. Borges Neto, Non-destructive fraud detection in rosehip oil by MIR spectroscopy and chemometrics, Food Chem. 209 (2016) 228–233, https://doi.org/10.1016/j.foodchem.2016.04.051.
[46] M. R. Baqueta, A. Coqueiro, P. H. Março, and P. Valderrama, "Multivariate classification for the direct determination of cup profile in coffee blends via handheld near-infrared spectroscopy," *Talanta*, vol. 222, no. August 2020, p. 121526, 2021, doi: 10.1016/j.talanta.2020.121526.
[47] M. Manley, Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials, Chem. Soc. Rev. 43 (24) (2014) 8200–8214, https://doi.org/10.1039/c4cs00062e.
[48] J.M. Amigo, C. Santos, Preprocessing of hyperspectral and multispectral images, Data Handl. Sci. Technol. 32 (2020) 37–53, https://doi.org/10.1016/B978-0-444-63977-6.00003-1.
[49] J.J. Workman, Interpretive spectroscopy for near infrared, Applied Spectroscopy Reviews 31 (3) (1996) 251–320, https://doi.org/10.1080/05704929608000571.
[50] B.G. Osborne, Near-infrared Spectroscopy in Food, Analysis. (2006), https://doi.org/10.1002/9780470027318.a1018.
[51] J.J. Workman, L. Weyer, Practical Guide to Interpretive Near-Infrared, Spectroscopy. (2007), https://doi.org/10.1201/9781420018318.
[52] P. Williams, J. Antoniszyn, M. Manley, Near Infrared Technology : Getting the best out of the light, 1st ed., African Sun Media, 2019.
[53] C. Zhang, et al., Genome diversification in phylogenetic lineages I and II of Listeria monocytogenes: Identification of segments unique to lineage II populations, J. Bacteriol. 185 (18) (2003) 5573–5584, https://doi.org/10.1128/JB.185.18.5573-5584.2003.
[54] F. Muchaamba, A.K. Eshwar, M.J.A. Stevens, R. Stephan, T. Tasara, Different Shades of Listeria monocytogenes: Strain, Serotype, and Lineage-Based Variability in Virulence and Stress Tolerance Profiles, Front. Microbiol. 12 (January) (2022) 1–23, https://doi.org/10.3389/fmicb.2021.792162.
[55] E.T. Sumrall, et al., Phage resistance at the cost of virulence: Listeria monocytogenes serovar 4b requires galactosylated teichoic acids for InlB-mediated invasion, PLoS Pathog. 15 (10) (2019) 1–29, https://doi.org/10.1371/journal.ppat.1008032.

[56] K.I. Kamisango, et al., Structures and biological activities of peptidoglycans of Listeria monocytogenes and Propionibacterium acnes, J. Biochem. 92 (1) (1982) 23–33, https://doi.org/10.1093/oxfordjournals.jbchem.a133918.
[57] T. Brauge, et al., Teichoic acid is the major polysaccharide present in the Listeria monocytogenes biofilm matrix, FEMS Microbiol. Lett. 363 (2) (2015) 1–7, https://doi.org/10.1093/femsle/fnv229.
[58] K. ichi Uchikawa, I. Sekiawa, I. Azuma, Structural studies on teichoic acids in cell walls of several serotypes of Listeria monocytogenes, J. Biochem. 99 (2) (1986) 315–327, https://doi.org/10.1093/oxfordjournals.jbchem.a135486.
[59] R. Capita, A. Felices-Mercado, C. Garci-Fernandez, and Alonso-Calleja, "Characterization of Listeria monocytogenes originating from the Spanish meat-processing chain," *Foods*, vol. 34, no. 8, 2019, [Online]. Available: www.mdpi.com/journal/foods.
[60] B.G. Osborne, T. Fearn, H. Hindle, Practical NIR spectroscopy with applications in food and beverage analysis, Longman Scientific & Technical 50 (1993).
[61] D.F. Barbin, G. Elmasry, D.W. Sun, P. Allen, N. Morsy, Non-destructive assessment of microbial contamination in porcine meat using NIR hyperspectral imaging, Innov. Food Sci. Emerg. Technol. 17 (2013) 180–191, https://doi.org/10.1016/j.ifset.2012.11.001.
[62] S. Brown, J.P. Santa Maria, S. Walker, Wall teichoic acids of gram-positive bacteria, Annu. Rev. Microbiol. 67 (2013) 313–336, https://doi.org/10.1146/annurev-micro-092412-155620.
[63] R.D. Abdi, J.R. Dunlap, B.E. Gillespie, D.B. Ensermu, R.A. Almeida, O. Kerro Dego, Comparison of Staphylococcus aureus surface protein extraction methods and immunogenicity, Heliyon 5 (10) (2019) e02528.
[64] T.J. Foster, J.A. Geoghegan, V.K. Ganesh, M. Höök, Adhesion, invasion and evasion: The many functions of the surface proteins of Staphylococcus aureus, Nat. Rev. Microbiol. 12 (1) (2014) 49–62, https://doi.org/10.1038/nrmicro3161.
[65] J.L. Xu, et al., Exploring the identification of multiple bacteria on stainless steel using multi-scale spectral imaging from microscopic to macroscopic, Sci. Rep. 12 (1) (2022) 1–21, https://doi.org/10.1038/s41598-022-19617-3.
[66] C. L. M. Morais *et al.*, *Standardization of complex biologically derived spectrochemical datasets*, vol. 14, no. 5. Springer US, 2019. doi: 10.1038/s41596-019-0150-x.
[67] M. Buxton, F. J. A. Van Ruitenbeek, M. Dalm, and M. W. N. Buxton, "Application of near-infrared (NIR) spectroscopy to sensor based sorting of an epithermal Au-Ag ore (revised version) Application of near-infrared (NIR) spectroscopy to sensor based sorting of an epithermal Au-Ag ore," no. March, 2015, doi: 10.13140/RG.2.1.1038.4489.
[68] E. Bonah, X. Huang, J.H. Aheto, R. Osae, Application of Hyperspectral Imaging as a Nondestructive Technique for Foodborne Pathogen Detection and Characterization, Foodborne Pathog. Dis. 16 (10) (2019) 712–722, https://doi.org/10.1089/fpd.2018.2617.
[69] E.L. Crowley, C.K. O'Sullivan, G.G. Guilbault, Increasing the sensitivity of Listeria monocytogenes assays: Evaluation using ELISA and amperometic detection, Analyst 124 (3) (1999) 295–299, https://doi.org/10.1039/a806875e.
[70] M. Farrés, S. Platikanov, S. Tsakovski, R. Tauler, Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation, J. Chemom. 29 (10) (2015) 528–536, https://doi.org/10.1002/cem.2736.
[71] E. Achata, C. Esquerre, C. O'Donnell, A. Gowen, A study on the application of near infrared hyperspectral chemical imaging for monitoring moisture content and water activity in low moisture systems, Molecules 20 (2) (2015) 2611–2621, https://doi.org/10.3390/molecules20022611.