



Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review



Dhritiman Saha, Annamalai Manickavasagan^{1,*}

School of Engineering, University of Guelph, N1G2W1, Canada

ARTICLE INFO

Keywords:

Food quality
Hyperspectral
Non-destructive testing
Machine learning
Deep learning
Classification

ABSTRACT

Non-destructive testing techniques have gained importance in monitoring food quality over the years. Hyperspectral imaging is one of the important non-destructive quality testing techniques which provides both spatial and spectral information. Advancement in machine learning techniques for rapid analysis with higher classification accuracy have improved the potential of using this technique for food applications. This paper provides an overview of the application of different machine learning techniques in analysis of hyperspectral images for determination of food quality. It covers the principle underlying hyperspectral imaging, the advantages, and the limitations of each machine learning technique. The machine learning techniques exhibited rapid analysis of hyperspectral images of food products with high accuracy thereby enabling robust classification or regression models. The selection of effective wavelengths from the hyperspectral data is of paramount importance since it greatly reduces the computational load and time which enhances the scope for real time applications. Due to the feature learning nature of deep learning, it is one of the most promising and powerful techniques for real time applications. However, the field of deep learning is relatively new and need further research for its full utilization. Similarly, lifelong machine learning paves the way for real time HSI applications but needs further research to incorporate the seasonal variations in food quality. Further, the research gaps in machine learning techniques for hyperspectral image analysis, and the prospects are discussed.

1. Introduction

Quality of food products plays a pivotal role in determining its preference for consumers, processors, and other stakeholders (Ali et al., 2020; Sharma et al., 2019). The testing of food quality has been largely subjective, laborious as well as destructive. Thus, many fast, reliable, and non-destructive techniques have been developed over the years for the determination of extrinsic and intrinsic quality parameters of food products. Imaging based non-destructive techniques, such as hyperspectral imaging (HSI), Raman imaging (RI), fluorescence imaging (FI), soft X-ray imaging, laser light backscattering, and magnetic resonance imaging (MRI) have become popular for food quality determination (Hussain et al., 2019).

Hyperspectral imaging (HSI) integrates the advantages of spectroscopy and imaging (spatial information) and has been successfully applied for the quantification of internal and external attributes of different food products (Mahesh et al., 2015). The limited abilities of the software and

hardware components of HSI system greatly reduces the rapid acquisition of image and its analysis. This limitation restricts the application of HSI system in on-line or real time industrial application.

One of the most important challenges with hyperspectral imaging is the extraction of useful information from the high dimensional hyperspectral data (hypercube) containing redundant information. Other challenges during hyperspectral imaging includes sensor noise, change in illumination and environmental factors, heterogeneity of sample and anisotropy. Hence, efficient algorithms and chemometrics are needed for reduction of dimensionality of hyperspectral data to improve the adaptation of HSI in real time food applications. The algorithms will help in reducing the computation time and process, improving the performance of the model and bring in robustness by reducing irrelevant variables and redundancies (Liu et al., 2017).

Machine learning grew as a subdomain of Artificial Intelligence (AI) that comprises algorithms capable of deriving useful information from data and utilizing that information in self-learning for making good

* Corresponding author.

E-mail addresses: dsaha@uoguelph.ca (D. Saha), mannamal@uoguelph.ca (A. Manickavasagan).

¹ Present address: Manickavasagan Annamalai, Associate Professor, Room 2401, Thornbrough Building, School of Engineering, University of Guelph, 50 Stone Road East, Guelph, Ontario, Canada, N1G 2W1 Tel.: +Phone: (519) 824 -4120x Ext: 53499; fax: +Fax: (519) 836 -0227, Email: mannamal@uoguelph.ca

<https://doi.org/10.1016/j.crfs.2021.01.002>

Received 7 December 2020; Received in revised form 15 January 2021; Accepted 26 January 2021

2665-9271/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

classification or prediction. Machine learning have gradually gained popularity due to its accuracy and reliability. Improved hardware and software components of machine vision systems have made the machine learning algorithms to process the data faster and give reliable decisions in very less time. Machine learning techniques have been widely applied in quality determination of agricultural and food products. Different machine learning techniques like Artificial Neural Network (ANN), Fuzzy logic, decision trees, Naïve Bayes, k-means clustering, support vector machines (SVM), random forest (RF), k-Nearest Neighbor (k-NN) and so on have been used extensively in agriculture related fields (Rehman et al., 2019). Deep learning is such another subdomain of machine learning that have shown superior performance in image classification of different food products and have established its potential to outperform even humans in some cases when trained adequately (Zhou et al., 2019).

Researchers have previously reviewed the different applications of machine learning techniques in food and agriculture field. However, there are no reviews exclusively on the application of different machine learning techniques in analysis of hyperspectral images of food materials. Hence, this review discusses the latest machine learning approaches used by researchers in analysis of hyperspectral data of food products, their characteristic features, and their performance while processing of hyperspectral images. The research gap, future trends and scope for development are also discussed and the authors feel that this work may act as a useful resource for the researchers working in the domain of hyperspectral imaging and machine learning applications in food products.

2. Hyperspectral imaging systems

In general, a hyperspectral system consists of a source of light, device for dispersion of wavelength, detector and a computer equipped with image acquisition software. The configuration mainly depends on the application type for which the HSI system is to be used for acquiring high quality hyperspectral images. In most cases of hyperspectral transmittance and reflectance imaging, tungsten halogen lamps are used as a source of light since it produces an uninterrupted spectrum in the visible to near infrared region and stable, durable and low cost. However, it has some disadvantages owing to which LED lights are gaining prominence now-a-days (Qin et al., 2013). The reflectance and transmission spectra of samples are captured using hyperspectral detectors. The most common types of detectors which are used in different spectral range in hyperspectral imaging are: Silicon: 3360–1050 nm; Lead Sulphide (PbS) – 1100–2500 nm; Indium–Gallium-Arsenide (InGaAs): 900–1700 nm. The quality of the images is generally determined through the performance of the detector. In general, a highly sensitive detector with elevated signal to noise ratio is preferred.

HSI systems operates in four modes based on the process of image acquisition mode viz., whiskbroom, staring, pushbroom and snapshot. Of the four modes, pushbroom hyperspectral imaging system, collecting spectra of line by line, is the most used for online applications in food industry (Jia et al., 2020). The snapshot technology is a non-scanning technique having no moving part and records a complete three-dimensional hypercube with each video frame. The snapshot technology is gaining importance with time since the image is captured entirely at once, but further research is required for more successful application of this HSI technology.

The most common measurement methods for hyperspectral imaging are reflectance, transmittance and interactance. The measurement technique is dependent upon sample type and the property being investigated. In general, reflectance mode is used in most agricultural applications since it can obtain relatively good useful information from the sample (El Masry et al., 2012). The analysis of the hyperspectral images is generally done through available softwares like Unscrambler, ENVI and so on, but these do not provide the options for online or real time image analysis. Over the past few decades, efficient machine learning techniques have been developed for the fast analysis of

hyperspectral images so that the entire system can be operated under online or real time conditions.

3. Machine learning techniques

Machine learning encompasses algorithms that possess the ability to learn from data without relying on explicit programming. It can be broadly classified into supervised, unsupervised and reinforcement learning. The different machine learning techniques are discussed in detail in the subsequent sections.

3.1. Supervised machine learning

Supervised learning requires learning a model from labelled training data that helps in making classification or prediction about the future data. Supervised indicates samples sets in which the desired output is known. In other words, the labelling of data is done to guide the machine to look for the exact desired pattern. Regression and classification are a subdomain of supervised learning (Garreta and Moncecchi, 2013). Some of the supervised learning tools are Artificial Neural Network, Decision Trees, Random Forest, Support Vector Machines k-Nearest Neighbor, Logistic Regression, Naïve Bayes and Linear Discriminant Analysis.

3.1.1. Artificial neural network (ANN)

Artificial neural networks (ANNs) were developed to imitate the functioning of human brain based on the working principle of biological neurons (Jamshidi, 2003). An ANN is a congregation of interconnected neurons having thresholds, weights and an activation function (Khaled et al., 2018). The simplest ANN is a multi-layer perceptron composed of an input layer, hidden layer, and output layer (Sanz et al., 2016). Neural networks have proven their effectiveness in pattern generation and classification whereby the feed forward neural networks have proven to be the most widely applied neural network (Nturambirwe and Opara, 2020). Back propagation is the method used for training the neural network. It involves the fine tuning of weights in a neural network based on the previous epochs (iteration) error rate.

ANNs have found its applications in detection of mechanical damage in mushrooms, single kernel wheat hardness estimation, cold injury in peaches, honey adulteration, prediction of firmness in kiwi fruit (Rojas-Moraleda et al., 2017; Erkinbaev et al., 2019; Pan et al., 2015; Shafiee et al., 2016; Siripatrawan et al., 2011). ANNs have been widely used as a single algorithm machine learning tool in hyperspectral image analysis (Table 1). The studies given in Table 1 involved the use of spectral pre-processing techniques like Savitzky-Golay first derivatives (SGD1), mean centering (MC), orthogonal signal correction (OSC) and multiplication scatter correction (MSC) for elimination of spectral noise and other non-useful spectral information. The spatial information from the hyperspectral imaging was segmented through global thresholding and Otsu algorithm to derive different features and required information about the region of interest. Once the spectral pre-processing and image processing is completed, use of Successive projections algorithm (SPA), ant colony optimization (ACO), Principal Component Analysis (PCA) was implemented for selection of key wavelengths and reduction of the redundant information generated through hyperspectral imaging (He et al., 2020). Most of the studies used back propagation (BP) feed forward artificial neural network for hyperspectral image analysis. At present, BP based classifiers are used extensively in different applications owing to its high classification accuracy, simplicity, robustness, sensitivity and automation (Golhani et al., 2018). Besides, the studies indicated that only three layers (input, hidden and output) with varying neurons in each layer is enough to build a model with high accuracy. Hence, addition of more hidden layers may slightly increase the accuracy but will also increase the computational load (Erkinbaev et al., 2019). The different type of transfer functions available for transfer of information from one layer to the other in a neural network are sigmoid function, linear transfer function, hyperbolic tangent function and logistic function. The studies

Table 1
Artificial Neural Network (ANN) applications in hyperspectral image analysis of food products.

Study	Wavelength range (nm)	Spectral pre-processing	Image processing	ANN characteristics								ANN Computational software	Classification accuracy	References
				Network type	Network topology						Training set: Validation set			
					Input layer		Hidden layer		Output layer					
					Number	Nodes	Number	Nodes	Number	Nodes				
Detection of mechanical damage in mushrooms	900–1700	Savitzky-Golay Second derivative	Harris corner detection algorithm	Polak–Ribie're conjugate gradient Back propagation	101	–	01	30	05	–	80:20	MATLAB R2012b	91%	Rojas-Moraleda et al. (2017)
Estimation of wheat hardness (single kernel)	1000–2500	Savitzky-Golay first derivatives (SGD1); mean centering (MC) and orthogonal signal correction (OSC)	Image thresholding	Two-layer Back Propagation neural Network (BPNN)	01	–	02	03	01	–	80:20	MATLAB 8.2	90%	Erkinbaev et al. (2019)
Detection of cold injury in peaches	400–1000	–	–	Back-propagation feed-forward neural network	01	420	01	03	01	02	80:20	–	96%	Pan et al. (2015)
Detection of adulteration in honey	400–1000	Savitzky-Golay algorithm (2nd-order polynomial with 3-point window)	Otsu algorithm for image thresholding	Back-propagation feed-forward neural network	01	–	01	10	01	–	70:30	MATLAB	95%	Shafiee et al., 2016
Detection of mites in flour	400–800	Multiplication scatter correction (MSC); Successive projections algorithm (SPA) and ant colony optimization (ACO) Normalization	Image thresholding	Back Propagation Neural Network	01	–	01	05	01	03	67:33	MATLAB R2017b	98%	He et al. (2020)
Detection of stored insects in rice and maize	400–1000	–	Otsu algorithm for image thresholding	Back Propagation Neural Network	01	–	03	–	01	–	60:40	MATLAB R2009b	98%	Cao et al. (2014)
Prediction of firmness in kiwi fruit	400–1000	Sawitzky–Golay algorithm with 2nd order polynomial	Image thresholding	Back-propagation feed-forward neural network	01	03	01	03	01	01	70:30	MATLAB	97%	Siripatrawan et al. (2011)
Detection of chilling injury in apple	400–1000	–	Global thresholding	Back-propagation feed-forward neural network	01	826	01	05	01	02	66:34	MATLAB 7.0	98.4%	Elmasry et al. (2009)
Differentiation of wheat classes	900–1700	–	Image cropping and statistical mean centering	BPNN Wardnet BPNN	01 01	75 75	01 01	79 78	01 01	08 08	60:40 for BPNN; 70:30 for Wardnet BPNN	MATLAB 7.0	90%	Mahesh et al. (2008)
Identification of wheat classes	900–1700	Normalization	Image cropping and thresholding	Back propagation neural network	01	100	01	–	01	08	60:40	MATLAB R2006a	92.1%	Choudhary et al. (2008)

discussed here involved the implementation of sigmoid function in transferring information from one layer to the other (Rojas-Moraleda et al., 2017). ANN behaves like a black box and hence it is important to properly tune the hyperparameters like learning rate, decay rate, momentum to reduce the chances of underfitting or overfitting (Cui et al., 2018). The values of the learning rate, momentum and initial weight used in these studies were 0.1, 0.1 and 0.3 respectively. The amount by which the weights are updated is referred to as the learning rate. Besides, momentum is responsible for accelerating the learning rate and decay rate is responsible for preventing the weights from growing too large. The studies highlighted that the classification accuracy or prediction of ANN have been more than 90% indicating its high efficiency in analysis of hyperspectral data. However, ANN requires a large data set for training to build a good model. It is imperative to say that both the spatial and spectral information obtained from hyperspectral imaging sometimes referred to as data fusion should be fully utilized for building a good and robust model.

3.1.1.1. Deep learning (DL). Deep learning is an effective machine learning algorithm used for extracting features from original data for classification, regression and detection. Deep learning involves a representation-learning method through utilization of the deep ANN comprising of multiple neuron layers. Convolutional neural network (CNN) is the most widely used class of deep neural network for analyzing images (Zhou et al., 2019). A typical CNN structure for classification problems consists of different layers namely convolutional layers, pooling layers and fully connected layers. The convolution layers consists of filters (kernels) with a specified stride and is responsible for the extraction of useful features such as edges, from the input data image. Stride can be explained as the number of shifts of pixels on the input data matrix. The pooling layer is responsible for reducing the spatial size of the input data thereby limiting the number of parameters and computation in the network. The pooling layer independently operates on each feature map, whereby max pooling being the most common approach used in the pooling process. In the fully connected layer, every node in the first layer is associated with every node in the second layer of the deep network system (Zhou et al., 2019).

The application of deep learning in hyperspectral imaging for food applications is relatively new. Some of its recent application in food include detection of bruises, diseases, identification of different varieties. In the studies discussed in Table 2, convolutional neural network (CNN) of deep learning architecture have been exclusively used for analysis of hyperspectral data. After performing spectral pre-processing and image segmentation, the depth features obtained from the hypercube is implicitly extracted by CNN. It can integrate the information between channels very efficiently in comparison to other traditional machine learning algorithms. However, there is need for further research to find the local correlation between image channels in hyperspectral imaging. This problem is also prevalent in computed tomography imaging (Wang et al., 2018). One of the most important aspects of deep learning is feature learning. In this context, a study by Zhang et al. (2020) highlighted that automatic feature extraction from raw hyperspectral data is achieved through a combination of 1D convolution and max pooling and a relationship was established between the features extracted and corresponding levels using fully connected block. Training samples determines the robustness and accuracy of the model. However, the improvement in model performance may not be significant enough after certain point due to redundant information in training samples of the hyperspectral data. Hence, there should be a trade off between performance and cost of model. For achieving higher model performance with reasonable cost, a hold-out test was suggested followed by gradual collection of samples till the test accuracy does not change significantly (Qiu et al., 2018). In some cases, the availability of abundant and reliable data is not available due to a number of factors. In such cases, some researchers like Weng et al. (2020) and Liu et al. (2019) have used principal

component analysis network (PCANet) and 2 Branch-CNN for building robust models using relatively small datasets. Deep learning can yield very poor performance when back propagation is directly applied in combination with gradient based optimization. Hence, studies suggested that a greedy layer-wise pretraining (training one layer at a time) can be used for improving the Stacked Sparse Auto-Encoder (SSAE) optimization (Liu et al., 2018). Qiu et al. (2018) found that the image patterns share common characteristics with the spectral curve patterns. In other words, the edges in images corresponds to minimum and peak in spectral curves. Yu et al. (2018) highlighted that the most significant information about the original input spectra is contained as deep spectral features in the last layer of the network. In most of the studies discussed, the kernel size of the convolutional layer has been taken as 3×3 with stride of 1 or 2 with the use of softmax function on the preceding fully connected layer with an average classification accuracy of more than 95%. Softmax function is generally applied to the immediately preceding fully connected layer in an image classification problem in CNN. The softmax yields the final output of the neural network and classifies an object having probabilistic values between 0 or 1. In CNN, the commonly used activation function is Rectified Linear Unit (ReLU). ReLU has the advantage of introducing non-linearity in the convolutional network and does not activate all neurons at the same time thereby reducing the computational load on the network. Qiu et al. (2018) reported the use of one padding layer in the convolutional neural network. It is the process of introducing additional layers of zeros to the input image that facilitates detailed representation of the information on the edges of the input image when kernel filters are applied. In one of the studies, Liu et al. (2019) applied batch normalization and dropout strategy technique in training deep neural networks for combating the problem of overfitting, reduction in the long training time and enhancement of accuracy of the model. Batch normalization involves standardizing the inputs to a layer for every mini batch. It enables higher learning rates and reduces the sensitivity to the weight initialization. Dropout strategy involves the process of ignoring the randomly selected neurons during training which helps the network to be less sensitive to specific weights of neurons thereby enabling better generalization. However, a study by Garbin et al. (2020) provided some suggestions in using the batch normalization and dropout strategy. The study indicated that the batch normalization technique generally improves the accuracy and should be given first preference for improving the convolutional neural networks whereas dropout strategy should be applied very carefully and not necessarily it may improve the accuracy each time.

3.1.2. Support vector machines (SVM)

Support Vector Machines (SVM) aims at obtaining the optimal hyperplanes (separating points of one class from the rest) through selection of ones passing through the largest gaps possible between points of different classes. New points are then classified to a certain class depending on the side of the surfaces they fall on. The process of creating an optimal hyperplane reduces the generalization error and thereby the chances of overfitting. The different kernel functions used in SVM are linear kernel function, radial basis function, polynomial kernel function and sigmoid kernel function. SVM is very effective while working with high dimensional spaces which require learning from several features in the problem. SVM has also been found to be effective when the data is relatively small i.e., a high dimensional space with few points (Raschka and Mirjalili, 2017). Besides, they require less memory storage as a subset of points is used only to represent the boundary surfaces. However, SVM models involve intensive calculations while the model is being trained. Further, they do not quantify the confidence percentage of a prediction which otherwise can be done through k-fold cross-validation with an increased computation cost.

The SVM based machine learning techniques have been mostly used in classification of different food products, agricultural crops, detection of diseases, adulteration, seed viability, quantification of chemical constituents in agricultural materials (Table 3). In the studies concerning the

Table 2
Deep Learning (DL) applications in hyperspectral image analysis of food products.

Study	Wavelength range (nm)	Spectral pre-processing	Image processing	Deep learning characteristics				DL Computation software	Classification accuracy	References
				Deep learning network type	Network topology and features	Parameter values/ Pertinent particulars	Training set: Validation set			
Detection of aflatoxin in peanut	400–1000	–	Image binarization and thresholding	Convolutional Neural Network (CNN)	1st layer- input; 2nd layer-convolution; 3rd layer-sub-sampling; 4th layer-convolution; 5th layer-sub-sampling; Output layer (fully connected); epochs: 1-100	Mean Error- 11.39–2.74%; Time required: 150–15000s	80:20	–	96%	Han and Gao (2019)
Detection of internal mechanical damage in blueberries	400–1000	–	Subsampling, image resizing, data augment and normalization	Two convolutional neural networks used: Residual Network (ResNet) and ResNeXt	Convolution layer filter size: 3×3 ; stride:2; Activation function: Rectified Linear Unit (ReLU)	Learning rate, decay rate and decay step: 0.1, 0.1 and 32,000	80:20	MATLAB R2014a	88%	Wang et al. (2018)
Determination of chemical compositions in dry black goji berries	900–1700	–	Image thresholding	Convolutional Neural Network (CNN)	One-dimension (1D) convolution layers, max pooling layers, ReLU activations, a fully connected layer. convolution kernel size: 3×3 ; stride:1; Max pooling layers: 2; stride:2	Learning rate and batch size: 0.005 and 5	65:35	MATLAB R2014b; PYTHON 3	88%	Zhang et al. (2020)
Determination of rice varieties	400–1000	Multivariate scatter correction (MSC), standard normal variate (SNV), Savitzky–Golay smoothing and Savitzky-Golay's first-order	Texture parameters calculation: gray-level gradient co- occurrence matrix (GLGCM), discrete wavelet transform (DWT) and Gaussian Markov random field (GMRF)	Principal component analysis network (PCANet) deep learning network	–	–	75:25	MATLAB R2017b; PYTHON	98.57	Weng et al. (2020)
Detection of internal defects in cucumber	400–1000	–	Image thresholding	Convolutional Neural Network -Stacked Sparse Auto-Encoder (CNN-SSAE) deep learning architecture	Greedy layer-wise unsupervised pretraining; Staking of additional output layer (softmax classifier) on pre-trained SSAE; training through gradient descent with back-propagation	Sparsity control parameter (β)-0.1 on encoding neurons; two layers: 16 encoding neurons in each layer	80:20	–	91%	Liu et al. (2018)
Prediction of firmness and soluble solid content of pear	400–1000	Multiplicative signal correction (MSC); successive projections algorithm (SPA)	Image thresholding	Stacked auto-encoders (SAE) and fully-connected neural network (FNN)	Feature extraction from hyperspectral data through SAE	Input for FNN: SAE extracted features	80:20	MATLAB 8.1; PYTHON	Firmness: 89%; Soluble solid content: 92%	Yu et al. (2018)
	900–1700	–	Wavelet transform (Daubechies 8- basis)	Convolutional neural network	Two convolutional layers, max pooling layer, fully	Learning rate- 0.0005	80:20	–	92%	Qiu et al. (2018)

(continued on next page)

Table 2 (continued)

Study	Wavelength range (nm)	Spectral pre-processing	Image processing	Deep learning characteristics				DL Computation software	Classification accuracy	References
				Deep learning network type	Network topology and features	Parameter values/ Pertinent particulars	Training set: Validation set			
Identification of rice variety (single seed)			function; decomposition level 3); Image thresholding	(CNN) adapted from Visual Geometry Group (VGG) Net	connected layer, dropout and dense layers (output layer). Kernel size-3x3; stride-1; padding-1; epochs:200; ReLU activation function; softmax function on output					
Detection and quantification of nitrogen content in rapeseed leaf	400–1000	–	Image thresholding	Stacked auto-encoders (SAE) and fully-connected neural network (FNN)	Feature extraction from hyperspectral data through SAE	Input for FNN: SAE extracted features	80:20	MATLAB 8.1; PYTHON	90%	Yu et al. (2019)
Classification of coffee bean varieties	900–1700	Savitzky–Golay first-order derivative	Image segmentation: watershed algorithm	Two branch convolutional neural network (2B–CNN)	1st branch: 1D convolution of spectral features; 2nd branch: 2D convolution of spatial features; Fully connected layer-Absent; Training- Batch normalization and dropout strategy; epochs: 100	Trained weights: effective wavelengths indicator	80:20	–	95%	Liu et al. (2019)
Detection of bruises in strawberry	900–1700	Savitzky–Golay first-order derivative	Image segmentation: watershed algorithm	Two branch convolutional neural network (2B–CNN)	1st branch: 1D convolution of spectral features; 2nd branch: 2D convolution of spatial features; Fully connected layer-Absent; Training- Batch normalization and dropout strategy; epochs: 200	Trained weights: effective wavelengths indicator	80:20	–	99%	Liu et al. (2019)

Table 3
Support Vector Machines (SVM) applications in hyperspectral image analysis of food products.

Study	Wavelength range (nm)	Spectral pre-processing	Image processing	SVM characteristics				SVM Computational software	Classification accuracy	References
				Kernel function	Parameter values/ Pertinent particulars	Cross validation	Training set: Validation set			
Detection of early decay in strawberry through prediction of total water-soluble solids	1000–2500	Standard normal variate correction (SNV); successive projection algorithm (SPA)	Image masking	Radial basis function	Gamma (γ): 1 Penalty factor (c): 3.16	Five-fold	70:30	MATLAB R2014	94%	Liu et al. (2019)
Detection and identification of fungal infection in cereals	400–1000	Successive projection algorithm (SPA)	Image cropping and thresholding	Radial basis function	Grid search optimization method: Kernel parameter values	Five-fold	67:33	–	99%	Lu et al. (2020)
Classification of foodborne bacterial pathogens grown on agar plates	400–1000	Standard Normal Variate (SNV); CARS (Competitive Adaptive Weighted sampling)	Image thresholding	Radial basis function	Optimization algorithm: Particle Swarm Optimization; Kernel parameter (γ): 46.20; Penalty factor (c): 1.45	Five-fold	70:30	MATLAB R2018a	98%	Bonah et al. (2020)
Classification of infected maize kernels	900–1700	Successive projection algorithm (SPA) Image cropping	Ostu segmentation and watershed algorithms	Radial basis function	Grid search optimization method: Kernel parameter values	Five-fold	70:30	MATLAB R2013b	100%	Chu et al. (2020)
Degree of aflatoxin contamination in peanut kernels	400–720	Fisher method: obtaining narrow band spectrum	De-noising, contrast enhancement; Image thresholding	Radial basis function	Grid search optimization method: Kernel parameter values	Five-fold	70:30	MATLAB R2015b	96%	Zhongzhi et al. (2020)
Detection of black spot disease in pear	400–1000	1st order derivative, multiplicative signal correction (MSC), and mean centering	Image segmentation: Spectral angle mapper	Radial basis function	–	Five-fold	70:30	MATLAB R2017a	98%	Pan et al. (2019)
Identification of adulterated cooked millet flour	900–1700	CARS (Competitive Adaptive Weighted sampling)	Image thresholding	Radial basis function	Grid search optimization method: Kernel parameter values	Ten-fold	67:33	MATLAB R2011b	100%	Shao et al. (2018)
Determination and visualization of soluble solids content in winter jujubes	Spectral range 1: 400–1000; Spectral range 2: 900–1700	Wavelet transform and moving average smoothing; area normalization; successive projection algorithm (SPA)	Image segmentation: mask creation	Radial basis function on LS-SVM	Regularization parameter (γ): 5.750×10^7 ; Kernel parameter (σ^2): 9.760×10^4	–	70:30	MATLAB R2017b	Spectral range 1: 89%; Spectral range 2: 87%	Zhao et al. (2020)
Classification of maize seed	400–1000	Normalization	Image segmentation: Adaptive threshold segmentation	Radial basis function	–	Ten-fold	50:50	MATLAB R2009b	94%	Xia et al. (2019)
Determination of viability of corn seed	1000–2500	Standard normal variate (SNV), Savitzky-Golay 2nd derivative	Image thresholding	Linear basis function	–	Ten-fold	70:30	–	100%	Wakholi et al. (2018)

application of SVM classifier in analysis of hyperspectral images, different spectral pre-processing techniques like standard normal variate correction (SNV), Savitzky-Golay derivatives, multiplicative signal correction (MSC) and mean centering were used for improving the spectral features whereas image segmentation techniques like Otsu algorithm, watershed algorithm, thresholding and spectral angle mapper were used for spatial feature extraction from the hyperspectral images. Besides, effective wavelength selection from the wavelength range was carried out using successive projection algorithm (SPA) and CARS (Competitive Adaptive Weighted sampling) for improving the classification model. In most of the studies (Table 3) conducted on the analysis of hyperspectral images, radial basis kernel function (RBF) of SVM was used. The radial basis function (RBF) is a non-linear function and it reduces the complexity of the training process (Lu et al., 2020). The tuning of two parameters namely regularization parameter/penalty factor (C) and kernel parameter (γ) is very vital since it helps in improving the accuracy level of the RBF based SVM classifier. The regularization parameter C is a trade-off between smooth decision boundary and correct classification of training points. Higher value of C promotes overfitting whereas lower values of C promotes underfitting. The gamma parameter in RBF regulate the influence of a single training example. Higher values of gamma indicate highly flexible or non-linear boundaries and low values of gamma indicate a more linear boundary. Most of the recent studies involved the application of grid search (GS) algorithm in optimization of kernel parameters for improving the classification accuracy (Table 3). However, GS algorithm involves higher computational time and works good only with low dimensional dataset having few parameters. In a study by Bonah et al. (2020), the concept of genetic algorithm (GA), and particle swarm optimization (PSO) was introduced for optimization of kernel parameters of SVM in improving the classification accuracy. Among the algorithms used, the PSO algorithm enhanced the classification accuracy of SVM to 100% for training set and 98.44% for prediction set. PSO has the advantage of preventing the data points of being trapped in local optima followed by increased accuracy and lower training time (Cho and Hoang, 2017). Bonah et al. (2020) also introduced the use of Least Square SVM (LS-SVM) in their studies instead of SVM. One of the limitations of SVM lies in constrained optimization programming which have been overcome by LS-SVM that applies linear equations instead of quadratic programming. LS-SVM has been found to have good prediction with faster execution time in comparison to SVM. While dealing with LS-SVM, two parameters namely regularization parameter-gamma (γ) and kernel parameter (σ^2) need proper tuning for yielding good results. The kernel parameter (σ^2) is also designated as squared bandwidth and if the value of this parameter is too low it leads to overfitting whereas an extremely higher value leads to underfitting to the sample data (Yasin et al., 2014). In another study, Chu et al. (2020) involved the use of object wise and pixel-wise approach for feature extraction from hyperspectral images for classification of infected maize kernels. It was observed that the pixel wise approach improved the classification accuracy of SVM to 100% in comparison to object wise approach. In object-wise approach, the average spectra of individual kernel are analyzed whereas in pixel-wise approach, the individual pixel in the region of interest is taken into consideration for analysis. Hence, the information obtained through pixel-wise approach is much more exhaustive than object-wise approach. Besides, pixel-wise classification helped in generating better visualization maps representing the spatial knowledge of the infected kernels. The studies involving the use of SVM classifier (Table 3) extensively used k-fold cross validation for model verification. In most cases, the value of k is either 5 or 10 indicating a 5-fold cross validation or 10-fold cross validation for verification of developed SVM model.

3.1.3. Decision trees (DT)

Decision trees represents a structure like a tree having internal nodes representing a test on a feature, each branch representing the result of a test, and each leaf node representing the class label followed by the

execution of decision after considering all the features. The classification rules in a decision tree represent a pathway from root to leaf. Hence, a decision tree comprises of three types of nodes: Root nodes, internal nodes and leaf nodes. It can handle different data types such as numeric, ratings, categorical and are also capable of handling missing data in response as well as independent variables. Decision tree is based on a series of Boolean tests. The working of a decision tree starts with the greedy algorithm, in which the tree is structured in a top-down iterative divide-and-rule approach. In the initial stage, the root node comprises of the training data set. The input data is divided iteratively based on selected features. The test features at each node are splitted based on decision tree functions like Gini index and entropy. The Gini index or impurity is a measure of a criterion to lessen the probability of misclassification. Entropy or information gain provides with the amount of disorder in a set which means that when entropy is zero, all the points of the target classes are the same. Several separate tree models can be combined to enhance the performance of a model better known as ensemble learning. The different decision tree algorithms used for model development are Iterative Dichotomizer 3 (ID3), C 4.5 and classification and regression tree (CART). One major advantage of a decision tree is the non-requirement of creation of dummy variables. However, the key issue with the decision tree is the large growth of the tree resulting in one leaf per observation. Besides, it is impossible to reconsider a decision once the training data set have been divided for answering a problem (Swamy-nathan, 2017).

The similarity between decision tree algorithm and human thinking process has led to its adoption in different fields like detection and identification of diseases in food products, evaluation of food quality, classification of agricultural products (Table 4). Different spectral pre-processing techniques like Savitzky-Golay derivatives, multiplicative signal correction (MSC) and normalization were used for improving the spectral features whereas image segmentation techniques like histogram thresholding, global thresholding and Gray-level co-occurrence matrix (GLCM) were used for extracting spatial information from the hyperspectral images. Sequential forward selection (SFS) was used for selection of effective wavelength for improving the model classification accuracy. Ren et al. (2020) in his studies used three different decision trees such as fine tree, medium tree and coarse tree for quality evaluation of black tea using hyperspectral data and concluded that the fine tree model outperformed the other two decision tree models. The selection of a tree and a subsequent good fit depends on the factor of minimum conflict of the tree with the training data. The present study used Gini index as split attribute criteria for choosing the optimal target and quality of the model. The authors also used the concept of data fusion (spectral and textural) for building robust classification models using hyperspectral data. In another study, Velásquez et al. (2017) reported the successful classification of fat and meat based on pixel level information of the hyperspectral data and provided an efficient way of classifying beef marbling. Gomez-Sanchis et al. (2008) used CART model of decision tree for decay classification in mandarin using hyperspectral data. The advantage with CART is that the presence of outliers does not affect the model, thereby paving the way for working with dimensionality data such as pixel classification of hyperspectral data. The other type of decision tree model used by researchers is Logistic Model Tree (LMT). LMT is a combination of logistic regression and C 4.5 decision tree learning method. Information gain is used for splitting and LogitBoost algorithm for creating logistic regression in each tree node followed by pruning through CART to eliminate the problem of overfitting (Chen et al., 2017; Luo et al., 2019). The authors (Ropelewska et al., 2018; Baranowski et al., 2013) reported high classification accuracy with hyperspectral data while using LMT model of decision tree. In most of the studies, it has been found that the classification accuracy is more than 90% which indicates the robustness of the decision tree as a classifier.

3.1.4. Random forest (RF)

A random forest can be imagined as a congregation of decision trees.

Table 4
Decision trees (DT) applications in hyperspectral image analysis of food products.

Study	Wavelength range (nm)	Spectral pre-processing	Image processing	Decision trees characteristics				DT Computational software	Classification accuracy	References
				Decision tree model	Parameter values/Pertinent particulars	Cross-validation	Training set: Validation set			
Evaluation of black tea quality	900–1700	Multiplicative scatter correction (MSC)	Texture parameters calculation: Gray-level co-occurrence matrix (GLCM)	Fine tree model	Split attribute criteria: Gini index; No of splits: 100	Five-fold	67:33	MATLAB R2017b	93%	Ren et al. (2020)
Classification of beef marbling	400–1000	–	Global thresholding	Classification and Regression Trees (CART)	–	Five-fold	71:29	MATLAB R2010a	99%	Velásquez et al. (2017)
Detection of codling moth infestation in apples	400–1000	Normalization; sequential forward selection (SFS)	Histogram thresholding	Classification and Regression Trees (CART)	–	Four-fold	80:20	MATLAB R2014b	82%	Rady et al. .2017
Detection of microbial spoilage in mushroom	400–1000	–	Image thresholding	Classification and Regression Trees (CART)	–	Five-fold	57:43	MATLAB 7.0	95%	Gaston et al. (2011)
Classification of infected and healthy wheat kernels	400–1000	–	Image thresholding	Logistic model tree (LMT)	–	Ten-fold	80:20	Waikato Environment for Knowledge Analysis (WEKA) 3.9	97%	Ropelewska et al. (2018)
Classification of bruised apples	400–1000	Savitzky–Golay method (second derivative)	Otsu thresholding algorithm	Logistic model tree (LMT)	Minimum number of instances:15; Number of Boosting Iterations: –1	Ten-fold	83:17	WEKA	98%	Baranowski et al. (2013)
Detection of Marssonina blotch in apples	400–1000	Savitzky–Golay method (second derivative)	Image thresholding	Classification and Regression Trees (CART)	Split attribute criteria: Gini index; No of splits: 100	Ten-fold	80:20	MATLAB R2014a	80%	Shuaibu et al. (2018)
Prediction of beef tenderness	400–1000	–	Image thresholding	Classification and Regression Trees (CART)	–	Five-fold	63:37	MATLAB	84%	Konda Naganathan et al., 2015
Classification of decay in mandarins	400–1000	–	Image thresholding	Classification and Regression Trees (CART)	–	Five-fold	80:20	–	93%	Sanchis et al., 2013
Identification of aflatoxin contaminated corn kernels	400–1000	–	Image thresholding	Classification and Regression Trees (CART)	–	Five-fold	50:50	MATLAB; WEKA	90%	Zhu et al. (2015)

Table 5
Random Forest (RF) applications in hyperspectral image analysis of food products.

Study	Wavelength range (nm)	Spectral pre-processing	Image processing	Random Forest characteristics		RF Computational software	Classification accuracy	References
				Number of decision trees	Training set: Validation set			
Detection of scab disease on potatoes	900–1700	–	Greedy Stepwise; Image thresholding; Otsu algorithm; Gaussian blurring cluster	500	75:25	WEKA	97%	Dacal-Nieto et al. (2011)
Detection of bruises in apple	400–1000	–	Image thresholding; Otsu algorithm	130	75:25	PYTHON	100%	Che et al. (2018)
Identification of rice seed cultivar	900–1700	–	Image thresholding	–	75:25	MATLAB R2009b	100%	Kong et al. (2013)
Classification of degree of bruising in apples	400–1000	Standard normal variate (SNV), 1st Derivative, Savitzky-Golay (SG) smoothing	Image thresholding	–	70:30	MATLAB 9.0; PYTHON	92%	Tan et al. (2018)
Determination of honey floral origin	400–1000	–	Image thresholding	–	70:30	MATLAB R2012a	92%	Minaei et al. (2017)
Inspection for varietal purity of rice seed	900–1700	–	Image thresholding	500	75:25	MATLAB	84%	Vu et al. (2016)
Identification of freezer burn on frozen salmon surface	900–1700	Standard normal variate (SNV)	Image thresholding	50	75:25	MATLAB R2015b	98%	Xu et al. (2016)
Detection and classification of virus on tobacco leaves	400–1000	Standard normal variate (SNV); Successive projections algorithm (SPA)	Image thresholding	71	67:33	MATLAB	85%	Zhu et al. (2017)
Discrimination of kiwifruits treated with different concentrations of forchlorfenuron	900–1700	Standard normal variate (SNV); Successive projections algorithm (SPA)	Image thresholding	200	67:33	MATLAB R2012a	94%	Dong et al. (2017)
Detection of fungal infection in strawberry	400–1000	Baseline correction; Savitzky-Golay second derivate	Image thresholding	10	75:25	WEKA	89%	Siedliska et al. (2018)

In random forests, a decision tree is created with a subset of training examples which are selected on random basis with replacement. Further, random number of features are also used at each set from the set of features. This process of tree growing is continued numerous times, thereby creating a set of classifiers. At time of prediction, in each instance, each grown tree predicts its target class in a similar way as done in decision trees. The class which is voted the most by the trees i.e., the class which is predicted most by the trees becomes the suggested one by the classifier. Random forest involves the averaging of multiple decision trees that suffer individually from high variance for building a more robust model having a better performance and less prone to overfitting. The pruning of random forest is non-essential as the classifier is quite strong to noise emanating from individual decision trees. The parameter which requires most attention is the number of trees to be chosen for the random forest. In general, more the number of trees, better the performance of the model or classifier obtained at the cost of higher computational cost. The effect of overfitting in random forest can be reduced by decreasing the size of the bootstrap samples which may increase the randomness of the random forest (Raschka and Mirjalili, 2017). However, reduction in the size of the bootstrap samples will negatively affect the overall performance of the random forest. In most practical applications, the bootstrap sample size is taken to be the same as the number of samples in the original training data set, thereby providing a good trade off between bias and variance. Random forests are not as easily interpretable as decision trees and the application of random forest is not favorable where the number of features are less (Garreta and Moncecchi, 2013).

Random forests have gained much importance during the last decade in its application in machine learning for their sound performance in

classification, ease of use and scalability. Random forest has been successfully applied for analysis of hyperspectral images for detection of plant diseases, fungal infection and bruises in fruits and vegetables, classification of different agricultural products, quality of processed fish products (Table 5). The different spectral pre-processing techniques applied in the studies were Savitzky-Golay derivatives, Standard Normal Variate (SNV) and baseline correction. Image processing techniques like Otsu thresholding and Gaussian blurring cluster were used for processing the spatial information. Besides, effective wavelength selection from the wavelength range was carried out using successive projection algorithm (SPA) for building a robust classification model. Che et al. (2018) in his study reported that random forest can combine weak classifiers to obtain a strong classifier having high classification accuracy with strong anti-noise ability. Since random forest is a combination of many decision trees, the finding of optimal decision trees is important for robust model development. In this study, the authors used exhaustive grid search method for finding an optimal number of decision trees with the highest classification accuracy. Dong et al. (2017) reported that the number of decision trees are randomly generated through bootstrap sampling (random replacement sampling) in which around two-third of the original samples are included in a bootstrap sample and the remaining one-third sample do not make any contribution, generally referred to as out-of-bag (OOB) samples. Bootstrap sampling decreases the correlation among the trees through introduction of different training sets. Xu et al. (2016) reported that the selection of optimal number of trees is obtained through careful inspection of the change in the out-of-bag error with accumulation of trees. In general, the use of more decision trees provides for a more robust estimate from out-of-bag (OOB) predictions. However, the cost and time involved in computation also increases which

Table 6
k-Nearest Neighbor (k-NN) applications in hyperspectral image analysis of food products.

Study	Wavelength range (nm)	Spectral pre-processing	Image processing	k-NN characteristics		k-NN Computational software	Classification accuracy	References
				Value of k	Training set: Validation set			
Assessment of packaged cod Classification of coffee species	400–1000	Area Normalization; 1st derivate	Image thresholding	3	75:25	R	100%	Washburn et al. (2017) Calvini et al. (2015)
	900–1700	Standard Normal Variate; 1st derivative; mean centering	Image thresholding	5	73:27	MATLAB 7.0	100%	
Detection of aflatoxin in maize	400–1000	Multiplicative signal correction (MSC)	Image thresholding	–	82:18	MATLAB R2018b	99%	Gao et al. (2020)
Detection of pesticide residue on spinach leaves	900–1700	Multiplicative signal correction (MSC)	Image thresholding	–	80:20	MATLAB R2016b; PYTHON 3.6	99%	Zhan-qi et al. (2018)
Classification of fat and lean tissue in packed salmon	400–1000	Mean-centering and unit variance normalization	Image thresholding	17	80:20	MATLAB R2012a	100%	Ivorra et al. (2016)
Evaluation of sugar content in different potato varieties	400–1000	Weighted baseline	Image thresholding	3 and 5	75:25	MATLAB 7.0	86%	Rady et al. (2015)
Classification of contaminants in wheat	900–1700	Standard Normal Variate (SNV)	Image thresholding, background separation	–	75:25	MATLAB 8.1	>90%	Ravikanth et al. (2015)
Classification of fresh Atlantic salmon fillets	400–1000	Standard Normal Variate (SNV)	Image thresholding	3	80:20	Interactive Data Language (IDL) 7.1	88%	Sone et al. (2011)
Classification of black beans	400–1000	Standard Normal Variate (SNV) Successive projections algorithm (SPA)	Textural attributes extraction: Gray level co-occurrence matrix	–	80:20	MATLAB 2009	98%	Sun et al. (2016)
Identification of states of wheat grain	900–1700	Standardization and multiple scattering correction	Image thresholding	7 (for reverse side of wheat grain); 6 (ventral side of wheat grain)	75:25	MATLAB R2018b	95%	Zhang et al. (2019)

ultimately decreases the performance of the model. Hence, there should be a trade-off between the number of trees and the performance of the model (Xu et al., 2016). In another study by Tan et al. (2018), random forest has been applied for development of bruise identification model in apples. The authors found that by applying image thresholding operation, it is very difficult to obtain a segmented image containing complete bruised area of an apple with high chances of misjudgment for the edge of the bruised area. Hence, supervised training of spectra for bruised and non-bruised area of the apple was done in random forest for building a good identification model. The developed model was able to successfully predict the transition between the non-bruised and bruised area of the apple, thereby enabling accurate identification of the bruised area through extraction of the spectral data. Vu et al. (2016) highlighted in his study that instead of using the average spectrum on all the pixels of the hypercube, the spectral data at each pixel may be more useful while investigating the chemical features of a food product. In his study, he found that when spectral and spatial features are combined for building the random forest model, the accuracy increased from 74% to 84%. Dacal-Nieto et al. (2011) in his study have provided a guide for selection of m_{try} parameter in building of random forest model. The parameter (m_{try}) is a measure of the number of variables available for splitting at each node of the tree. In this study, the value of m_{try} was selected as the square root of p which represents the number of features of the problem. The studies involving application of random forest in analysis of hyperspectral data showed high level of accuracy (>90%) in most cases.

3.1.5. k -nearest neighbor (k-NN)

k-Nearest Neighbor (k-NN) involves the storing of all available cases followed by classification of new cases based on a similarity metric i.e., distance. k-NN usually uses three distance metrics namely Manhattan distance (city block), Euclidean distance (Frobenius), P norm distance (Minkowsky) while calculating the distance between the points. Based on the distance metric selected, the k-NN algorithm searches the training dataset for the k samples that are nearest to the point to be classified. The new data point is assigned a class label of the new data point through majority vote among its k nearest neighbors. The optimal value of k is important for finding a good balance between underfitting and overfitting. If the value of k is too small, it will be more prone to noise points and if the k value is too large, the neighbourhood may comprise of points from other classes. The main advantages of k-NN is that the cost of the learning process is nil. No optimization is required and its easy to program with high accuracy (Raschka and Mirjalili, 2017). It is worthwhile to report that k-NN is very prone to overfitting due to the curse of dimensionality. The curse of dimensionality describes a situation in which the feature space tends to become increasingly scattered for a higher number of dimensions for a training dataset of fixed size. In other words, the closest neighbors being relatively far away in a high-dimensional space can give a very good estimate (Garreta and Moncecchi, 2013).

The simplicity and high accuracy of k-NN machine learning algorithm has encouraged its applications in different areas such as classification of different food varieties, detection of toxic substances and contaminants

in food products, pesticide residue in leafy vegetables, chemical constituents based classification of food products (Table 6). In this ML algorithm, the performance of the model is mainly influenced by the number of nearest neighbors (Washburn et al., 2017). The studies conducted involved the implementation of different spectral pre-processing and image processing techniques to the raw hyperspectral data. Thereafter, the authors varied the value of k to observe a possible improvement in the performance of the model. Most of the studies have experimented with the value of k from 3 to 6. The values of k are generally not chosen to be either 1 or 2 owing to the mechanism of tie-breaker in k -NN (Borges et al., 2015). As a rule of thumb, the choice of k is obtained by the square root of number of samples under study (Duda and Hart, 1973). Among the three-distance metric used in k -NN algorithm, the Euclidean distance is preferred by researchers since it results in good classification accuracy (Zhang et al., 2019). k -NN has not only found its applications as a good ML algorithm but also used by researchers in optimal wavelength (Xin et al., 2019) and feature selection (Zhan-qi et al., 2018) of hyperspectral data. Xin et al. (2019) proposed a fast and reliable method of obtaining best wavelet decomposition layer and effective wavelength selection of hyperspectral data through coupling of wavelet basis functions and k -nearest neighbor. The different wavelet basis functions like db6, sym5, sym7 and db4 were used for selection of effective wavelength of hyperspectral data. The decomposition of the original spectral signal was carried out through wavelet transform followed by use of k -NN algorithm for analyzing high frequency signal of each layer of wavelet decomposition leading to the selection of best layer and effective wavelengths of hyperspectral data. Zhan-qi et al., 2018 used Chi-square test for feature selection of hyperspectral data combined with k -NN in detection of pesticide residue on spinach leaves. The Chi-square test calculates and rank the correlation degree between each dimension category and features and ultimately retains the most related dimensional features. After the Chi-square feature selection process, the prediction accuracy obtained through k -NN algorithm in the study was found to be 99%. In a study by Guo et al. (2018), higher accuracy of hyperspectral image classification is obtained when k -NN is combined with guided filter. The authors used joint representation k -NN with front and posterior guided filter for extracting spatial information and performing denoising operation respectively to obtain high classification accuracy. Most of the studies indicated high accuracy (>90%) while applying k -NN for analysis of hyperspectral data.

3.1.6. Logistic regression (LR)

Logistic Regression belongs to the class of supervised learning and is used in classification problems. Basically, it is based on the concept of probability and is a predictive analysis algorithm. In general, Logistic regression is used for binary classification of materials. It results in a discrete binary outcome between 0 and 1. Logistic Regression evaluates the relationship between the independent variable (features) and the dependent variable (label, to be predicted) by calculating the probabilities using the logistic function (Swamynathan, 2017). The difference between linear regression and logistic regression lies in the fact that the outcome of the logistic regression is discrete whereas linear regression yields a continuous value. It is widely used owing to its high efficiency, simple, less computation and is highly interpretable. However, non-linear problems cannot be solved with logistic regression since it provides linear decision surface. It is mostly used when the data is linearly separable. Logistic Regression can predict only a categorical outcome and is prone to overfitting.

The application of logistic regression has primarily been carried out for classification of land cover using hyperspectral data in remote sensing (Gewali et al., 2018). However, in one of the studies by Sanz et al. (2016), classification of lamb muscle based on hyperspectral data was carried out using Logistic regression. It was reported that a classification accuracy of

92% was achieved when principal component analysis (PCA) was combined with logistic regression. In this study, the authors highlighted that the parameters (θ) of the model require learning through the training set and then the probability is calculated for classifying a new example. It has reported that the accuracy obtained in other classification problems using hyperspectral data was relatively low (Wang et al., 2018). Hence, in logistic regression, efficient pre-processing of hyperspectral data is very important before it is being fed to the LR classifier.

3.1.7. Naïve Bayes (NB)

Naïve Bayes is a powerful yet simple generative machine learning classifier that utilizes the concept of conditional probability (Bayes' theorem) to describe the outcome probabilities of related events. In other words, it evaluates the probability of an instance belonging to a class based on the probabilities value of each of the feature. The naïve word highlights the assumption that each feature is independent and identically distributed than the others indicating that a feature value has no relationship with the value of another feature (Rehman et al., 2019). At first, a frequency table (similar to prior probabilities) of all classes is created by the algorithm followed by the creation of a likelihood table. Thereafter, the posterior probability is calculated. In general, Naïve Bayes has three models namely Multinomial model, Poisson model and Bernoulli model. Due to its simplicity, it has been used in many domains with high accuracy. The major drawback of this algorithm is that it is incapable of learning the interaction between two predictor variables/features due to the assumption of conditional independence (Bishop, 2006).

Naïve Bayes have been successfully used on hyperspectral data in bruise and cultivar detection of apples (Siedliska et al., 2014), detection of contaminants in wheat (Ravikanth et al., 2015), detection of chilling injury in cucumber (Cen et al., 2016). The classification accuracy in the above hyperspectral imaging studies have been found to be more than 85%, 95% and 98% respectively. Most of the researchers have used the multinomial naïve Bayes classifier model in their studies with hyperspectral data (Zhang et al., 2020). However, in several studies, it has also been reported that the classification accuracy of the naïve Bayes classifier is less than 75%, indicating a weak classification model (Qin et al., 2020; Siedliska et al., 2017, 2018).

3.1.8. Linear discriminant analysis (LDA)

LDA is a linear transformation method that reduces the number of dimensions in a dataset. LDA is considered as a supervised learning algorithm, hence considered to have better feature extraction techniques than Principal Component Analysis (PCA). The principle underlying LDA involves finding the feature subspace that optimizes separability of class. One of the assumptions in LDA is the normal distribution of the data and statistical independence of the features. However, LDA can work still reasonably well if the assumptions are violated (Raschka and Mirjalili, 2017). Linear Discriminant Analysis (LDA) is generally used for feature extraction and aids in increasing the computational efficiency, reduction in the degree of overfitting due to the curse of dimensionality models that are non-regularized. It highlights the accuracy of the classification.

Due to the robustness of the LDA, it has been widely applied for classification of agricultural and food products based on hyperspectral data (Qin et al., 2020; Delwiche et al., 2019; Liu et al., 2010; Mahesh et al., 2008). In most of the studies applying LDA for classification using hyperspectral imaging, the average classification accuracy has been reported to be more than 90% indicating the robustness of the classifier. In a recent study, Xia et al. (2019) used the concept of multi-linear discriminant analysis (MLDA) as a feature transformation in his studies concerning identification of different maize varieties using hyperspectral imaging. MLDA can be described as an improvement and extension of LDA which provides for a multi-linear projection and mapping of the input data from one space to another. MLDA based classification model

obtained a significantly better average classification accuracy (99.13%) than LDA based classification model (90.13%).

3.2. Unsupervised machine learning

Unsupervised learning involves dealing with unlabeled data or unknown data structure. It explores the data structure to obtain meaningful information without the help of a known outcome variable. Clustering and dimensionality reduction are a subcategory of unsupervised learning (Raschka and Mirjalili, 2017). Some of the unsupervised learning tools are k-means clustering, Independent Component Analysis (ICA), Principle Component Analysis (PCA).

3.2.1. k-means clustering

k-means algorithm involves organizing data into clusters with the aim of achieving high similarity between intra-cluster and low similarity between inter-cluster. An item can belong to only one cluster since it produces a definite number of non-hierarchical and disjoint clusters. k-means is an instance of the expectation maximization (EM) algorithm and applies an iterative way of minimizing the intra-cluster Sum of Squared Errors (SSE). The initial step begins with selection of randomly picked centroids designated by the symbol k. Centroids can be explained as the average location or arithmetic mean of all the points. The points which are closest to each centroid point are assigned to that specific cluster. Now, the centroid is recalculated by averaging the position coordinates of all the points present in that cluster. The process is continued until the convergence of the clusters take place (Garreta and Moncecchi, 2013). In general, the distance between the centroid and the points is calculated by Euclidean distance metric. k-means algorithm facilitates easy implementation when compared with other clustering algorithms. However, k-means clustering requires the declaration of the number of clusters. k-means clustering faces issues when clusters are of different size, non-globular shapes and densities. Further, the occurrence of outlier can lead to misrepresentation of the results.

The simplicity and the computational speed have encouraged the use of k-means clustering in unsupervised classification using hyperspectral imaging. Liu et al. (2010) applied k-means clustering for classification of pork samples using hyperspectral data. In this study, Gabor filtering was used for preprocessing of hyperspectral images followed by k-means clustering. In k-means clustering, the authors used three distance metrics namely city-block distance, Euclidean distance and cosine distance for calculating the distance between the points and centroid. It was reported that the use of cosine distance in k-means clustering algorithm achieved the highest accuracy of 83%. In another study by Liu et al. (2012), an accuracy of 100% was reported in classification of eggs into fertile and non-fertile ones on the 0th day of incubation and 84% accuracy on 4th day of incubation. Singh et al. (2007) used k-means clustering for detection of fungal infection in wheat but the performance of the classifier was found to be poor (accuracy < 70%) in comparison to other discriminant classifiers. k-means clustering has been applied for good segmentation of potato hyperspectral images in non-destructive detection of potato quality. The excellent segmentation provided by k-means clustering helped in improving the classification accuracy of the discriminant models for potato quality determination (Ji et al., 2019).

3.2.2. Dimensionality reduction

3.2.2.1. Principal component analysis (PCA).

Principal Component Analysis is about the creation of new set of uncorrelated variables from a set of possibly correlated variables. The newly created variables lie in a new coordinate system where the projected data in the first coordinate represents the highest variance followed by the projected data in the second coordinate representing the second highest variance and so on. The newly formed coordinates are named as the principal components (PC). Besides, transforming the original data, the PCA provides with two parameters namely Eigen vectors and Eigen values that help in the

interpretation of the chemometric analysis. The Eigen values gives the information about the variation in each PCA band and the Eigen vectors or loading vectors gives the information about the weighting function for obtaining the PCA scores. In general, the number of principal components obtained is equal to the number of original dimensions, but those PCs having higher variance are selected. The condition of orthogonality (uncorrelated) is to be complied with remaining principal components when each new principal component is added. When the first principal components are retained, the reduction in the dimensionality of data occurs, thus helping in visualization of the data. Similarly, when the first and second principle components are retained, the examination of data can be done using a two-dimensional scatter plot (Raschka and Mirjalili, 2017). As a result, PCA can be used as a resource tool for exploratory data analysis before creating predictive models. The major advantage of PCA lies in obtaining a low-dimensional space from a high-dimensional one while retaining variance to the maximum extent possible and hence protect the model from the curse of dimensionality. PCA does not require a ground truth in performing its projections; it only depends on the learning features values.

The popularity and the powerful nature of PCA makes it the most widely used dimensionality reduction techniques in building robust classification models using hyperspectral data (Rojas-Moraleda et al., 2017; Erkinbaev et al., 2019; He et al., 2020; Vu et al., 2016; Zhu et al., 2016; Dong et al., 2017). Konda Naganathan et al. (2015) used different PCA techniques namely chemometric principal component analysis (CPCA), sample principal component analysis (SPCA) and mosaic principal component analysis (MPCA) in reducing the spectral dimensionality of the hyperspectral images of beef. It was reported that CPCA with first five loading vectors gave the best performance than the other two PCA techniques. Besides, the CPCA operates on one-dimensional data through averaging of spatial pixels of hyperspectral images and hence require minimum time in creating the loading vectors. The studies conducted using PCA on hyperspectral data have been limited to use of a maximum of first five loading vectors or score images by the researchers. All the relevant and useful features of interest have been found within this first five score images. Besides, in most cases, it has been reported that the application of PCA have significantly improved the classification accuracy of the ML classifier.

3.2.2.2. Independent component analysis (ICA).

Independent component analysis (ICA) is considered as a further step of principal component analysis (PCA) technique and considered as a powerful tool for extraction of source signals or useful information from the original data. PCA follows the principle for optimization of covariance matrix of the data representing statistics of second-order, whereas ICA follows the optimization of statistics of higher-order such as kurtosis. Hence, it can be said that while PCA obtains uncorrelated components, ICA yields independent components (Raschka and Mirjalili, 2017). The extraction of the independent components is done by a) Non-Gaussianity maximization, b) mutual information minimization, or c) using maximum likelihood (ML) estimation method. ICA have been used in image segmentation for extraction of different useful layers from the original image.

The applications of Independent component analysis have been extensively used in signal processing arena (Tharwat, 2018). The studies concerning the application of ICA in processing of spectral data reported that the original spectra are decomposed into source signals which ultimately simplifies the interpretation of the results (Chuang et al., 2014; Boiret et al., 2014). However, limited studies have used ICA for processing of hyperspectral images. One noteworthy study was performed by Mishra et al. (2019) for detection of peanut flour in wheat flour using ICA and hyperspectral imaging. To achieve this, the authors used Random ICA by blocks and Joint Approximation Diagonalization of Eigen-matrices (JADE) algorithm to obtain the optimal number of independent components that represents the source signals of different chemical constituents from the original data set. The study concluded

that the optimal number of independent components obtained as seven which was sufficient for determination of peanut traces distribution in the sample using the hyperspectral images.

3.3. Reinforcement machine learning

Machine learning has been broadly classified into unsupervised learning and supervised learning. However, reinforcement learning (RL) refers to the application of specific task-oriented algorithms in a manner to learn achieving a complex objective (goal) or the way to maximize along a particular dimension over many steps (Suuton and Barto, 1998). RL mimics the way humans and animals learn in the absence of a mentor. RL is based on the concept of interaction of the agent with the environment. For example, humans and animals learn to walk in the absence of a mentor whereas agents will acquire the learning by trial and error. RL is based on giving rewards, to agents when they complete the assigned work by themselves (Aljaafreh, 2017). RL problems are mostly modeled as a Markov decision process (MDP). A Markov decision process generally involves a 5-tuple (S, A, P, R, γ), where S represents a finite set of states, A represents a finite actions set, P represents the probability of transition, R represents the reward provided immediately, and γ represents the factor of discount. Reinforcement learning application may transform the scenario of automation in agriculture and food industry as it can be used for teaching robots to adjust their temporal behavior according to the relation between them and the surroundings (Bechar and Vigneault, 2016).

4. Research gap

The industrial application of hyperspectral imaging (HSI) for quality inspection of food products is challenged with different limitations it possesses. Most of the on-going research in this field is limited to laboratory-scale. There are challenges exist in abilities of the hardware and software of the HSI systems. The extraction of useful information from the high dimensional hyperspectral data is a cumbersome task. The cost of the hyperspectral system is very high which limits its application in real world. Most of the machine learning algorithms used for analysis of hyperspectral data involves manual feature extraction which greatly increases the computation time. In addition to the spectral data, the spatial information obtained from hyperspectral data need to be utilized to the maximum extent i.e. fusion of spectral and spatial data for developing more robust models. The review of literature revealed repeated use of only specific machine learning algorithms in analysis of hyperspectral images. The studies related to application of deep learning algorithms in food products is limited and requires further research for its full utilization. At present, the popular ML algorithms work in isolation i.e., it executes a ML algorithm based on a training dataset to develop a model and hence the ML algorithm does not make any effort in retaining the knowledge learned and use them in future learning. The studies which have been incorporated in this review mainly highlights the application of machine learning in horticultural crops whereas comparatively less studies are available on cereals, pulses and oilseeds. Though there have been studies related to hyperspectral detection of insect affected grains, pulses, very few studies have reported the application of machine learning for sorting or grading based on quality characteristics (both external and internal) of these crops.

5. Future trends and scope for development

The establishment of hyperspectral imaging system in the food industry depends largely on the successful address of the different issues hindering its applications. Future work should involve the way for minimizing the cost of the hyperspectral imaging device through development of low-cost materials for fabrication. With the advancement in computing system, improved hardware and software of the HSI system should be developed for rapid processing of hyperspectral images.

Currently, the machine learning algorithms has a narrow specific application in food which requires standardization for wider applications (Nturambirwe and Opara, 2020).

The use of advanced machine learning algorithms like deep learning and life-long machine learning should be used more effectively for its potential in real time applications. Deep learning employs automatic feature learning from the hyperspectral data unlike other traditional machine learning algorithms. The identification of effective wavelength (EW) regions from the entire working spectrum in hyperspectral imaging is important for real time online applications since EW selection reduces the equipment cost and computational load of HSI. In this context, a pioneering work has been reported by Liu et al. (2019) in selection of effective wavelengths and spectral-spatial classification in hyperspectral imaging. It was found that two branch convolutional neural network (2B-CNN) based on deep learning has excellent accuracy and involved less computation time thereby facilitating for real time online applications. But more research is required for improving the robustness of the proposed 2B-CNN model. Very few studies have indicated the potential of deep learning models for building prediction models. Further, research is required for developing effective deep learning models for prediction which can outperform other prediction development methods. In deep learning process, a considerable time is taken for the training process followed by high complexity and several hyperparameters of the model which complicates the optimization process (Zhou et al., 2019). Further, deep learning involves training of huge amount of data for good classification accuracy. So, more research in this direction is required for development of simpler networks based on deep learning approach.

Environmental factors play an important role in the dynamic change of the quality pattern in food products over time. Hence, the potential of incremental learning or lifelong machine learning approach may be utilized for building models with high classification or prediction accuracy. Lifelong learning (LL) involves a reinforcement learning approach and use of the accumulated knowledge over time in future learning and solving problems. With time, the LL algorithm gains more and more knowledge and becomes more efficient in learning. The continuous learning ability of LL algorithm mimics the human intelligence system (Chen and Liu, 2018).

In hyperspectral imaging system, most of the research work involved the analysis of the spectral information. However, the spatial information obtained through hyperspectral imaging is not fully utilized and can provide some key information (Han et al., 2019). The combination of the spectral and spatial information (at pixel level) for image processing will help to achieve a better classification accuracy of the developed model. The relationship between the spectral dimension and classification accuracy need to be analyzed and an effective weighted filter may be designed for good classification of hyperspectral images (Guo et al., 2018).

The potential of Independent Component Analysis (ICA) in dimensionality reduction of hyperspectral data has not been utilized fully. Future studies on hyperspectral imaging should involve the use of ICA in studies concerning detection of adulterants in different food products. It is very difficult to choose an algorithm that will solve most of the problems, hence choosing an appropriate algorithm is very important for effectiveness of the model. Hence, future work is required to build a framework for algorithms that can be recommended for specific applications (Nturambirwe and Opara, 2020).

6. Conclusions

Hyperspectral imaging technique is a powerful tool for non-destructive assessment of quality in agricultural products. However, the huge amount of information generated by HSI is difficult to process and that limits its use in real time industrial applications. Furthermore, extracting useful information from the high dimensional hyperspectral data containing redundant information is a challenging task. Hence, for making online hyperspectral imaging inspection a reality, emerging and efficient algorithms are needed. In this context, machine learning

algorithms can play an effective role in analysis of hyperspectral images with high accuracy. Besides, advanced machine learning algorithms like deep learning have found its potential application in hyperspectral image analysis of agricultural products. Since deep learning involves automatic feature learning during the training stage, it has more potential for real time applications than other traditional machine learning algorithms. The scope of lifelong machine learning should be explored further, and its application should be extended to other agricultural crops for quality monitoring. More future work is required in developing simpler networks based on deep learning and lifelong learning for reducing the high complexity and optimization task involved in implementing these advanced ML algorithms for analysis of hyperspectral images.

CRedit authorship contribution statement

Dhritiman Saha: Formal analysis, Writing - original draft, Literature collection, Data analysis, Manuscript writing. **Annamalai Manickavasagan:** Writing - review & editing, Formal analysis, Reviewing manuscript, Data Analysis, Financial support for this study.

Declaration of competing interest

1. This article does not contain any studies with human or animal subjects
2. The outcome of this research is not associated with funding organizations.

Acknowledgement

Financial support provided by Indian Council of Agricultural Research (ICAR), India [ICAR-IF 2018-19, F. No. 18(01)/2018-EQR/Edn] is gratefully acknowledged. The authors are grateful for the funding from NSERC (Discovery Grant), Canada and Barrett Family Foundation, Canada.

References

- Ali, M.M., Hashim, N., Abd Aziz, S., Lasekan, O., 2020. Principles and recent advances in electronic nose for quality inspection of agricultural and food products. *Trends Food Sci. Technol.* <https://doi.org/10.1016/j.tifs.2020.02.028>.
- Aljaafreh, A., 2017. Agitation and mixing processes automation using current sensing and reinforcement learning. *J. Food Eng.* 203, 53–57.
- Bechar, A., Vigneault, C., 2016. Agricultural robots for field operations: concepts and components. *Biosyst. Eng.* 149, 94–111.
- Baranowski, P., Mazurek, W., Pastuszka-Wó, J., 2013. Supervised classification of bruised apples with respect to the time after bruising on the basis of hyperspectral imaging data. *Postharvest Biol. Technol.* 86, 249–258.
- Bonah, E., Huang, X., Yi, R., Harrington Aheto, J., Yu, S., 2020. 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.* <https://doi.org/10.1016/j.infrared.2020.103220>.
- Borges, E.M., Lafayette, J.M., Gelinski, N., Cristina De Oliveira Souza, V., Barbosa, F., Lemos Batista, B., 2015. Monitoring the authenticity of organic rice via chemometric analysis of elemental data. *Food Res. Int.* 77, 299–309.
- Boiret, M., Rutledge, D.N., Gorretta, N., Ginot, Y.-M., Roger, J.M., 2014. Application of independent component analysis on Raman images of a pharmaceutical drug product: pure spectra determination and spatial distribution of constituents. *J. Pharmaceut. Biomed. Anal.* 90, 78–84.
- Calvini, R., Ulrici, A., Amigo, J.M., 2015. Practical comparison of sparse methods for classification of Arabica and Robusta coffee species using near infrared hyperspectral imaging. *Chemometr. Intell. Lab. Syst.* 146, 503–511.
- Cao, Y., Zhang, C., Chen, Q., Li, Y., Qi, S., Tian, L., Ren, Y., 2014. Identification of species and geographical strains of *Sitophilus oryzae* and *Sitophilus zeamais* using the visible/near-infrared hyperspectral imaging technique. *Pest Manag. Sci.* 71, 1113–1121.
- Che, W., Sun, L., Zhang, Q., Tan, W., Ye, D., Zhang, D., Liu, Y., 2018. Pixel based bruise region extraction of apple using Vis-NIR hyperspectral imaging. *Comput. Electron. Agric.* 146, 12–21.
- Chen, Z., Liu, B., 2018. Lifelong machine learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 12, 1–207.
- Choudhary, R., Mahesh, S., Paliwal, J., Jayas, D.S., 2008. Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. *Biosyst. Eng.* 102, 115–127.
- Chu, X., Wang, W., Ni, X., Li, C., Li, Y., 2020. Classifying maize kernels naturally infected by fungi using near-infrared hyperspectral imaging. *Infrared Phys. Technol.* <https://doi.org/10.1016/j.infrared.2020.103242>.
- Cho, M.Y., Hoang, T.T., 2017. Feature selection and parameters optimization of svm using particle swarm optimization for fault classification in power distribution systems. *Comput. Intell. Neurosci.* 2017, 4135465, 9 pages. <https://doi.org/10.1155/2017/4135465>.
- Chuang, Y.K., Hu, Y.P., Yang, I.C., Delwiche, S.R., Lo, Y.M., Tsai, C.Y., Chen, S., 2014. Integration of independent component analysis with near infrared spectroscopy for evaluation of rice freshness. *J. Cereal. Sci.* 60, 238–242.
- Cen, H., Lu, R., Zhu, Q., Mendoza, F., 2016. Nondestructive detection of chilling injury in cucumber fruit using hyperspectral imaging with feature selection and supervised classification. *Postharvest Biol. Technol.* 111, 354–361.
- Chen, W., 2017. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* 151, 147–160.
- Cui, S., Ling, P., Zhu, H., Keener, H.M., 2018. Plant pest detection using an artificial nose system: a review. *Sensors* 18, 1–18.
- Dacal-Nieto, A., Formella, A., Carrión, P., Vazquez-Fernandez, E., Fernández-Delgado, M., 2011, September. Common scab detection on potatoes using an infrared hyperspectral imaging system. In: *International Conference on Image Analysis and Processing*. Springer, Berlin, Heidelberg, pp. 303–312.
- Duda, R.O., Hart, P.E., 1973. *Pattern Classification and Scene Analysis*, vol. 3. Wiley, New York, pp. 731–739.
- Delwiche, S.R., Torres Rodriguez, I., Rausch, S.R., Graybosch, R.A., 2019. Estimating percentages of fusarium-damaged kernels in hard wheat by near-infrared hyperspectral imaging. *J. Cereal. Sci.* 87, 18–24.
- Dong, J., Guo, W., Zhao, F., Liu, D., 2017. Discrimination of hayward kiwifruits treated with forchlorfenuron at different concentrations using hyperspectral imaging technology. *Food Analytical Methods* 10, 477–486.
- Elmasry, G., Wang, N., Vigneault, C., 2009. Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks. *Postharvest Biol. Technol.* 52, 1–8.
- Erkinbaev, C., Derksen, K., Paliwal, J., 2019. Single kernel wheat hardness estimation using near infrared hyperspectral imaging. *Infrared Phys. Technol.* 98, 250–255.
- ElMasry, G., Barbin, D.F., Sun, D.W., Allen, P., 2012. Meat quality evaluation by hyperspectral imaging technique, an overview. *Crit. Rev. Food Sci. Nutr.* 52, 689–711.
- Garbin, C., Zhu, X., Marques, O., 2020. Dropout vs. batch normalization: an empirical study of their impact to deep learning. *Multimed. Tool. Appl.* 79, 12777–12815.
- Gao, J., Ni, J., Wang, D., Deng, L., Li, J., Han, Z., 2020. Pixel-level aflatoxin detecting in maize based on feature selection and hyperspectral imaging. *Spectrochim. Acta Mol. Biomol. Spectrosc.* <https://doi.org/10.1016/j.saa.2020.118269>.
- Gaston, E., Frias, J.M., Cullen, P., Gaston, E., Cullen, J., 2011. Hyperspectral imaging for the detection of microbial spoilage of mushrooms. In: *Oral Presentation at the Meeting of 11th International Conference of Engineering and Food*, Athens, Greece.
- Gewali, U.B., Monteiro, S.T., Saber, E., 2018. Machine Learning Based Hyperspectral Image Analysis: A Survey arXiv preprint arXiv:1802.08701.
- Golhani, K., Balasundram, S.K., Vadmalai, G., Pradhan, B., 2018. A review of neural networks in plant disease detection using hyperspectral data. In: *Information Processing in Agriculture*, vol. 5. China Agricultural University, pp. 354–371. Issue 3.
- Gómez-Sanchis, J., Blasco, J., Soria-Olivas, E., Lorente, D., Escandell-Montero, P., Martínez-Martínez, J.M., Martínez-Sober, M., Aleixos, N., 2013. Hyperspectral LIGTFB-based system for classification of decay in mandarins caused by *Penicillium digitatum* and *Penicillium italicum* using the most relevant bands and non-linear classifiers. *Postharvest Biol. Technol.* 82, 76–86.
- Guo, Y., Han, S., Li, Y., Zhang, C., Bai, Y., 2018. K-Nearest Neighbor combined with guided filter for hyperspectral image classification. *Procedia Computer Science* 129, 159–165.
- Garreta, R., Moncecchi, G., 2013. *Learning Scikit-learn: Machine Learning in Python*. Packt Publishing Ltd, Birmingham, UK.
- Han, Z., Gao, J., 2019. Pixel-level aflatoxin detecting based on deep learning and hyperspectral imaging. *Comput. Electron. Agric.* <https://doi.org/10.1016/j.compag.2019.104888>.
- Hussain, N., Sun, D.W., Pu, H., 2019. Classical and emerging non-destructive technologies for safety and quality evaluation of cereals: a review of recent applications. *Trends Food Sci. Technol.* 91, 598–608.
- He, P., Wu, Yi, Wang, J., Ren, Yi, Ahmad, W., Liu, R., Ouyang, Q., Jiang, Hui, Chen, Quansheng, 2020. Detection of mites *Tyrophagus putrescentiae* and *Cheyletus eruditus* in flour using hyperspectral imaging system coupled with chemometrics. *J. Food Process. Eng.* <https://doi.org/10.1111/jfpe.13386>.
- Ivorra, E., Sanchez, A.J., Verdú, S., Barat, J.M., Grau, R., 2016. Shelf life prediction of expired vacuum-packed chilled smoked salmon based on a KNN tissue segmentation method using hyperspectral images. *J. Food Eng.* 178, 110–116.
- Ji, Y., Sun, L., Li, Y., Ye, D., 2019. Detection of bruised potatoes using hyperspectral imaging technique based on discrete wavelet transform. *Infrared Phys. Technol.* <https://doi.org/10.1016/j.infrared.2019.103054>.
- Jia, B., Wang, W., Ni, X., Lawrence, K.C., Zhuang, H., Yoon, S.C., Gao, Z., 2020. Essential processing methods of hyperspectral images of agricultural and food products. *Chemometr. Intell. Lab. Syst.* <https://doi.org/10.1016/j.chemolab.2020.103936>.
- Jamshidi, M., 2003. Tools for intelligent control: Fuzzy controllers, neural networks and genetic algorithms. *Phil. Trans. Math. Phys. Eng. Sci.* 361, 1781–1808.
- Konda Naganathan, G., Cluff, K., Samal, A., Calkins, C.R., Jones, D.D., Meyer, G.E., Subbiah, J., 2015. Three dimensional chemometric analyses of hyperspectral images for beef tenderness forecasting. *J. Food Eng.* 169, 309–320.

- Khaled, A.Y., Abd Aziz, S., Bejo, S.K., Nawi, N.M., Abu Seman, I., 2018. Spectral features selection and classification of oil palm leaves infected by Basal stem rot (BSR) disease using dielectric spectroscopy. *Comput. Electron. Agric.* 144, 297–309.
- Kong, W., Zhang, C., Liu, F., Nie, P., He, Y., 2013. Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis. *Sensors* 13, 8916–88927.
- Liu, L., Ngadi, M.O., 2012. Detecting fertility and early embryo development of chicken eggs using near-infrared hyperspectral imaging. *Food Bioprocess Technol.* 6, 2503–2513.
- Liu, L., Ngadi, M.O., Prasher, S.O., Gariépy, C., 2010. Categorization of pork quality using Gabor filter-based hyperspectral imaging technology. *J. Food Eng.* 99, 284–293.
- Liu, Y., Zhou, S., Han, W., Liu, W., Qiu, Z., Li, C., 2019. Convolutional neural network for hyperspectral data analysis and effective wavelengths selection. *Anal. Chim. Acta* 1086, 46–54.
- Liu, Yuwei, Pu, H., Sun, D.W., 2017. Hyperspectral imaging technique for evaluating food quality and safety during various processes: a review of recent applications. *Trends Food Sci. Technol.* 69, 25–35.
- Liu, Z., He, Y., Cen, H., Lu, R., 2018. Deep feature representation with stacked sparse auto-encoder and convolutional neural network for hyperspectral imaging-based detection of cucumber defects. *Transactions of the ASABE* 61, 425–436.
- Lu, Y., Wang, W., Huang, M., Ni, X., Chu, X., Li, C., 2020. Evaluation and classification of five cereal fungi on culture medium using Visible/Near-Infrared (Vis/NIR) hyperspectral imaging. *Infrared Phys. Technol.* <https://doi.org/10.1016/j.infrared.2020.103206>.
- Luo, X., Lin, F., Chen, Y., 2019. Coupling logistic model tree and random subspace to predict the landslide susceptibility areas with considering the uncertainty of environmental features. *Sci. Rep.* 9, 1–13.
- Mahesh, S., Jayas, D.S., Paliwal, J., White, N.D.G., 2015. Hyperspectral imaging to classify and monitor quality of agricultural materials. *J. Stored Prod. Res.* 61, 17–26.
- Mahesh, S., Manickavasagan, A., Jayas, D.S., Paliwal, J., White, N.D.G., 2008. Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes. *Biosyst. Eng.* 101, 50–57.
- Minai, S., Shafiee, S., Polder, G., Moghadam-Charkari, N., Van Ruth, S., Barzegar, M., Zahiri, J., Alewijn, M., Kusd, P.M., Kusd, K., 2017. VIS/NIR imaging application for honey floral origin determination. *Infrared Phys. Technol.* 86, 218–225.
- Mishra, P., Karami, A., Nordon, A., Rutledge, D.N., Roger, J.M., 2019. Automatic denoising of close-range hyperspectral images with a wavelength-specific shearlet-based image noise reduction method. *Sensor. Actuator. B Chem.* 281, 1034–1044.
- Nturambirwe, J.F.I., Opara, U.L., 2020. Machine learning applications to non-destructive defect detection in horticultural products. *Biosyst. Eng.* 189, 60–83.
- Pan, T.T., Chyngyz, E., Sun, D.W., Paliwal, J., Pu, H., 2019. Pathogenetic process monitoring and early detection of pear black spot disease caused by *Alternaria alternata* using hyperspectral imaging. *Postharvest Biol. Technol.* 154, 96–104.
- Pan, L., Zhang, Q., Zhang, W., Sun, Y., Hu, P., Tu, K., 2015. Detection of cold injury in peaches by hyperspectral reflectance imaging and artificial neural network. *Food Chem.* 192, 134–141.
- Qin, J., Vasefi, F., Hellberg, R.S., Akhbardeh, A., Isaacs, R.B., Yilmaz, A.G., et al., 2020. Detection of fish fillet substitution and mislabeling using multimode hyperspectral imaging techniques. *Food Contr.* <https://doi.org/10.1016/j.foodcont.2020.107234>.
- Qin, J., Chao, K., Kim, M.S., Lu, R., Burks, T.F., 2013. Hyperspectral and multispectral imaging for evaluating food safety and quality: a review. *J. Food Eng.* 118, 157–171.
- Qiu, Z., Chen, J., Zhao, Y., Zhu, S., He, Y., Zhang, C., 2018. Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network. *Appl. Sci.* 8, 212.
- Rady, A., Ekramirad, N., Adedeji, A.A., Li, M., Alimardani, R., 2017. Hyperspectral imaging for detection of codling moth infestation in GoldRush apples. *Postharvest Biol. Technol.* 129, 37–44.
- Rady, Ahmed, Guyer, D., Lu, R., 2015. Evaluation of sugar content of potatoes using hyperspectral imaging. *Food Bioprocess Technol.* 8, 995–1010.
- Ravikanth, L., Singh, C.B., Jayas, D.S., White, N.D.G., 2015. Classification of contaminants from wheat using near-infrared hyperspectral imaging. *Biosyst. Eng.* 135, 73–86.
- Ren, G., Wang, Y., Ning, J., Zhang, Z., 2020. Using near-infrared hyperspectral imaging with multiple decision tree methods to delineate black tea quality. *Spectrochim. Acta Mol. Biomol. Spectrosc.* <https://doi.org/10.1016/j.saa.2020.118407>.
- Rojas-Moraleda, R., Valous, N.A., Gowen, Aoié, Esquerre, Carlos, Härtel, Steffen, Salinas, Luis, O'donnell, Colm, 2017. A frame-based ANN for classification of hyperspectral images: assessment of mechanical damage in mushrooms. *Neural Comput. Appl.* 28, 969–981.
- Ropelewski, E., Zapotoczny, P., 2018. Classification of Fusarium-infected and healthy wheat kernels based on features from hyperspectral images and flatbed scanner images: a comparative analysis. *Eur. Food Res. Technol.* 244, 1453–1462.
- Rehman, T.U., Mahmud, M.S., Chang, Y.K., Jin, J., Shin, J., 2019. Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. *Comput. Electron. Agric.* 156, 585–605.
- Raschka, S., Mirjalili, V., 2017. *Python Machine Learning*. Packt Publishing Ltd, Birmingham, UK.
- Sanz, J.A., Fernandes, A.M., Barrenechea, E., Silva, S., Santos, V., Gonçalves, N., Paternain, D., Jurio, A., Melo-Pinto, P., 2016. Lamb muscle discrimination using hyperspectral imaging: comparison of various machine learning algorithms. *J. Food Eng.* 174, 92–100.
- Swamyathan, M., 2017. *Mastering Machine Learning with Python in Six Steps*. Apress Media, LLC, California.
- Sutton, R.S., Barto, A.G., 1998. *Introduction to Reinforcement Learning*, vol. 135. MIT press, Cambridge.
- Shao, Y., Xuan, G., Hu, Z., Gao, X., 2018. Identification of adulterated cooked millet flour with Hyperspectral Imaging Analysis. *IFAC-PapersOnLine* 51, 96–101.
- Sharma, S., Dhalsamant, K., Tripathy, P.P., 2019. Application of computer vision technique for physical quality monitoring of turmeric slices during direct solar drying. *Journal of Food Measurement and Characterization* 13, 545–558.
- Shuaibu, M., Lee, W.S., Schueller, J., Gader, P., Hong, Y.K., Kim, S., 2018. Unsupervised hyperspectral band selection for apple Marssonina blotch detection. *Comput. Electron. Agric.* 148, 45–53. <https://doi.org/10.1016/j.compag.2017.09.038>.
- Siedliska, A., Baranowski, P., Mazurek, W., 2014. Classification models of bruise and cultivar detection on the basis of hyperspectral imaging data. *Comput. Electron. Agric.* 106, 66–74.
- Siedliska, A., Baranowski, P., Zubik, M., Mazurek, W., 2017. Detection of pits in fresh and frozen cherries using a hyperspectral system in transmittance mode. *J. Food Eng.* 215, 61–71.
- Siedliska, A., Baranowski, P., Zubik, M., Mazurek, W., Sosnowska, B., 2018. Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging. *Postharvest Biol. Technol.* 139, 115–126.
- Singh, C.B., Jayas, D.S., Paliwal, J., White, N.D.G., 2007. Fungal detection in wheat using near-infrared hyperspectral imaging. *Transactions of the ASABE* 50, 2171–2176.
- Siripatrawan, U., Makino, Y., Kawagoe, Y., Oshita, S., 2011. Rapid detection of *Escherichia coli* contamination in packaged fresh spinach using hyperspectral imaging. *Talanta* 85, 276–281.
- Sone, I., Olsen, R.L., Sivertsen, A.H., Eilertsen, G., Heia, K., 2012. Classification of fresh Atlantic salmon (*Salmo salar* L.) filets stored under different atmospheres by hyperspectral imaging. *J. Food Eng.* 109, 482–489.
- Sun, J., Jiang, S., Mao, H., Wu, X., Li, Q., 2016. Classification of black beans using visible and near infrared hyperspectral imaging. *Int. J. Food Prop.* 19, 1687–1695.
- Tharwat, A., 2018. Independent component analysis: an introduction. *Applied Computing and Informatics*. <https://doi.org/10.1016/j.aci.2018.08.006>.
- Tan, W., Sun, L., Yang, F., Che, W., Ye, D., Zhang, D., Zou, B., 2018. Study on bruising degree classification of apples using hyperspectral imaging and GS-SVM. *Optik* 154, 581–592.
- Velásquez, L., Cruz-Tirado, J.P., Siche, R., Quevedo, R., 2017. An application based on the decision tree to classify the marbling of beef by hyperspectral imaging. *Meat Sci.* 133, 43–50.
- Vu, H., Tachtatzis, C., Murray, P., Harle, D., Dao, T.K., Le, T.L., et al., 2016. (November). Spatial and spectral features utilization on a Hyperspectral imaging system for rice seed varietal purity inspection. In: 2016 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future (RIVF). IEEE, pp. 169–174.
- Wakholi, C., Kandpal, M., Lee, H., Bae, H., Park, E., Kim, M.S., Mo, C., Lee, W.-H., Cho, B.K., 2018. Rapid assessment of corn seed viability using short wave infrared line-scan hyperspectral imaging and chemometrics. *Sensor. Actuator. B* 255, 498–507.
- Wang, Z., Hu, M., Zhai, G., 2018. Application of deep learning architectures for accurate and rapid detection of internal mechanical damage of blueberry using hyperspectral transmittance data. *Sensors* 18, 1126.
- Washburn, K.E., Kristian Stormo, S., Skjelvareid, M.H., Heia, K., 2017. Non-invasive assessment of packaged cod freeze-thaw history by hyperspectral imaging. *J. Food Eng.* 205, 64–73.
- Weng, S., Tang, P., Yuan, H., Guo, B., Yu, S., Huang, L., Xu, C., 2020. Hyperspectral imaging for accurate determination of rice variety using a deep learning network with multi-feature fusion. *Spectrochim. Acta Mol. Biomol. Spectrosc.* <https://doi.org/10.1016/j.saa.2020.118237>.
- Xia, C., Yang, S., Huang, M., Zhu, Q., Guo, Y., Qin, J., 2019. Maize seed classification using hyperspectral image coupled with multi-linear discriminant analysis. *Infrared Phys. Technol.* <https://doi.org/10.1016/j.infrared.2019.103077>.
- Xin, Z., Jun, S., Xiaohong, W., Bing, L., Ning, Y., Chunxia, D., 2019. Research on moldy tea feature classification based on WKN algorithm and NIR hyperspectral imaging. *Spectrochim. Acta Mol. Biomol. Spectrosc.* 206, 378–383.
- Xu, J.L., Sun, D.W., 2016. Identification of freezer burn on frozen salmon surface using hyperspectral imaging and computer vision combined with machine learning algorithm. *Int. J. Refrig.* 74, 151–164.
- Yasin, Z.M., Rahman, T.K.A., Zakaria, Z., 2014, December. Optimal least squares support vector machines parameter selection in predicting the output of distributed generation. In: 2014 2nd International Conference on Electrical, Electronics and System Engineering (ICEESE). IEEE, pp. 152–157.
- Yu, X., Lu, H., Wu, D., 2018. Development of deep learning method for predicting firmness and soluble solid content of postharvest Korla fragrant pear using Vis/NIR hyperspectral reflectance imaging. *Postharvest Biol. Technol.* 141, 39–49.
- Yu, X., Wang, J., Wen, S., Yang, J., Zhang, F., 2019. A deep learning based feature extraction method on hyperspectral images for nondestructive prediction of TVB-N content in Pacific white shrimp (*Litopenaeus vannamei*). *Biosyst. Eng.* 178, 244–255.
- Zhan-qi, R.E.N., Zhen-hong, R.A.O., Hai-yan, J.I., 2018. Identification of different concentrations pesticide residues of dimethoate on spinach leaves by hyperspectral image technology. *IFAC-PapersOnLine* 51, 758–763.
- Zhang, L., Ji, H., 2019. Identification of wheat grain in different states based on hyperspectral imaging technology. *Spectrosc. Lett.* 52, 356–366.
- Zhao, Y., Zhang, C., Zhu, S., Li, Y., He, Y., Liu, F., 2020. Shape induced reflectance correction for non-destructive determination and visualization of soluble solids content in winter jujubes using hyperspectral imaging in two different spectral ranges. *Postharvest Biol. Technol.* <https://doi.org/10.1016/j.postharvbio.2019.111080>.
- Zhongzhi, H., Limiao, D., 2020. Aflatoxin contaminated degree detection by hyperspectral data using band index. *Food Chem. Toxicol.* <https://doi.org/10.1016/j.fct.2020.111159>.
- Zhu, F., Yao, H., Hruska, Z., Kincaid, R., Brown, R.L., Bhatnagar, D., Cleveland, T.E., 2015. Visible near-infrared (VNIR) reflectance hyperspectral imagery for identifying

- aflatoxin-contaminated corn kernels. In: 2015 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers, p. 1.
- Zhu, H., Chu, B., Zhang, C., Liu, F., Jiang, L., He, Y., 2017. Hyperspectral imaging for presymptomatic detection of tobacco disease with successive projections algorithm and machine-learning classifiers. *Sci. Rep.* 7, 1–12.
- Zhang, C., Wu, W., Zhou, L., Cheng, H., Ye, X., He, Y., 2020a. Developing deep learning based regression approaches for determination of chemical compositions in dry black goji berries (*Lycium ruthenicum* Murr.) using near-infrared hyperspectral imaging. *Food Chem.* <https://doi.org/10.1016/j.foodchem.2020.126536>.
- Zhang, M., Jiang, Y., Li, C., Yang, F., Li, C., 2020b. Fully convolutional networks for blueberry bruising and calyx segmentation using hyperspectral transmittance imaging. *Biosyst. Eng.* <https://doi.org/10.1016/j.biosystemseng.2020.01.018>.
- Zhou, L., Zhang, C., Liu, F., Qiu, Z., He, Y., 2019. Application of deep learning in food: a review. *Compr. Rev. Food Sci. Food Saf.* 18, 1793–1811.