



Convolutional neural network ensemble learning for hyperspectral imaging-based blackberry fruit ripeness detection in uncontrolled farm environment

Chollette C. Olisah^{a,*}, Ben Trehwella^b, Bo Li^c, Melvyn L. Smith^a, Benjamin Winstone^a, E. Charles Whitfield^d, Felicidad Fernández Fernández^d, Harriet Duncalfe^e

^a University of the West of England (UWE), Bristol, UK

^b Opposable Games Ltd, UK

^c Syngenta, Jealott's Hill Research Centre, UK

^d NIAB, East Malling, UK

^e Berry Gardens Growers, UK

ARTICLE INFO

Keywords:

Blackberry ripeness
Convolutional neural network
Hyperspectral imaging
Homogeneous ensemble learning
Stack generalization ensemble
Uncontrolled farm environment

ABSTRACT

Fruit ripeness estimation models have for decades depended on spectral index features or colour-based features, such as mean, standard deviation, skewness, colour moments, and/or histograms for learning traits of fruit ripeness. Recently, few studies have explored the use of deep learning techniques to extract features from images of fruits with visible ripeness cues. However, the blackberry (*Rubus fruticosus*) fruit does not show obvious and reliable visible traits of ripeness when mature and therefore poses great difficulty to fruit pickers. The mature blackberry, to the human eye, is black before, during, and post-ripening. To address this problem, this paper proposes a novel multi-input convolutional neural network (CNN) ensemble classifier (MCE) for detecting subtle traits of ripeness in blackberry fruits. The multi-input CNN was created from a pre-trained visual geometry group 16-layer deep convolutional network (VGG16) model trained on the ImageNet dataset. The fully connected layers were optimized for learning traits of ripeness of mature blackberry fruits. The resulting model of about 600 K trainable parameters served as the base for building homogeneous ensemble learners t ensemble using the stack generalization ensemble (SGE) framework. The input to the network are images acquired with a stereo sensor using visible and near-infrared (Vis-NIR) spectral filters at wavelengths of 700 nm and 770 nm. Through experiments, the proposed model achieved 95.1 % accuracy on unseen sets and 90.2 % accuracy with in-field conditions. Further experiments reveal that machine sensory is highly and positively correlated to human sensory over blackberry fruit skin texture

1. Introduction

The blackberry fruit belongs to the class of soft fruits which is native to Europe, Asia, North, and South America.¹ The word 'soft' or 'hard' in relation to fruits implies the storage requirements and shelf life of the fruit. Soft² fruit examples, such as raspberry, blackberry, strawberry, and blueberry, usually require delicate handling and have a short shelf life compared to hard fruits² like banana, apple, mango, or dragon fruit. It is also worth noting that soft fruits are generally considered non-climacteric or ethylene insensitive, so it is not possible to pick them

unripe and schedule their ripening after a period of storage like it is possible with hard fruits. In 2018, in the United Kingdom alone, the market value of soft fruits was estimated at £670 million,² which indicates that there is an established and growing number of soft fruit consumers. Demand for soft fruit has risen, likely in part due to the promotion of eating healthy foods and the benefits of eating fruits rich in antioxidants (Zanini et al., 2015), and consumers expect excellent taste and quality. Typically, consumers associate the taste and quality of soft fruits with ripeness. Soft fruits, particularly blackberries, are best enjoyed when perfectly ripe. However, due to changes in growing

* Corresponding author.

E-mail address: chollette.olisah@uwe.ac (C.C. Olisah).

¹ <https://www.ecowatch.com/fruit-foragers-companion-2588139252.html>.

² <https://the-secret-garden.net/soft-fruit-vs-hard-fruit-what-are-the-differences-and-which-is-easier-to-grow.html>.

conditions and seasons, and the ripeness traits hidden in a mature blackberry (Mikulic-Petkovsek et al., 2021), it is difficult to visually differentiate between a mature ripe berry from a mature unripe berry at the point of picking from the plant. During the picking process, pickers need to make the decision on the berry to pick within approximately 3 s by evaluating the colour appearance and the size of the drupelet. This impacts growers' ability to meet market needs and expectations. To measure the ripeness and overall quality of blackberries, many growers carry out sampling and targeted assessments on their harvested crop, testing samples for firmness, sweetness, and flavour (Cockerton et al., 2020), and/or a biochemical analysis to measure the contents of sugar and organic acids (Mikulic-Petkovsek et al., 2021). The former is based on subjective sensory evaluation, while the latter is based on objective laboratory assays, which are more accurate but quite labour-intensive. Both types of measures can only evaluate a small subset of the harvested crops and are destructive, thus adding to food waste.

Fruit ripeness estimation using non-destructive techniques is a well-researched area that has advanced from pure computational processes (Polder et al., 2002; Sinelli et al., 2008; Li et al., 2018; Beghi et al., 2013; Shah et al., 2020) to processes enhanced by intelligence. For example, (El-Bendary et al., 2015; Hamza and Chtourou, 2018; Sabzi et al., 2019). Other examples with ready tools for in-field usage are (Cho et al., 2020; Gutiérrez et al., 2019; Simonyan and Zisserman, 2014).

Computational processes of ripeness estimation include methods that investigate and analyse fruit ripeness using Vis-NIR imaging and have utilized spectral indices (features) from fruit images to estimate ripeness. This has been the line of study in fruit ripeness assessment for many years now. Polder et al. (2002) acquired images at 257 spectral bands which were averaged to obtain wavelengths of 396 nm and 736 nm corresponding to the ripeness stages of tomatoes. Using spectral features to train a linear discriminant analysis (LDA) model, they were able to show that spectral features reduced the classification error from 51 % (with RGB features) to 19 %. Sinelli et al. (2008) utilized near-infrared (NIR) spectroscopy to examine blueberry ripeness. Using partial least squares (PLS) regression algorithms, they showed that spectral features correlated with ripeness. Li et al. (2018) used two hyperspectral cameras covering the wavelength range between 600 and 1700 nm and applied multivariate statistical analysis to predict the fruit quality attributes including, firmness, colour, and soluble solid content which achieved good performance in fruit ripeness estimation. Beghi et al. (2013) acquired blueberry images in field settings using spectral bands from 445 nm to 970 nm. Using PCA, two spectral wavelengths, 680 nm, and 740 nm were identified to correlate with ripe and near ripeness, respectively. The spectral features were further used to create a blueberry ripeness index which was integrated into the design of a micro-controller-based handheld optical device for assessing ripeness level. Several other studies can be found in the review work by Shah et al. (2020).

Ripeness estimation enhanced by artificial intelligence includes the fruit ripeness process researchers have approached by combining Vis-NIR or colour space conversion methods with machine learning. Here the discussion of the literature will progress from ripeness estimation of hard fruits to soft fruits. El-Bendary et al. (2015) converted RGB images of tomato fruit to the HSV colour space and extracted moment and histogram features which were further combined to form features, for discriminating between different levels of ripeness in tomato fruit. Using PCA, these features were downsampled and used as input to a support vector machine (SVM) classifier. The model achieved 84.30 % and 90.80 % on one-against-all ripeness levels and one-against-one ripeness levels, respectively. Hamza and Chtourou (2018) after applying a segmentation algorithm to separate apple fruit from the background image, extracted mean, variance, chromaticity, standard deviation, and skewness features which were trained using a multi-layered perceptron (MLP). Their method achieved a 98.33 % accuracy on the test set. Sabzi et al. (2019) also applied a segmentation algorithm to separate foreground apples from background images but extracted colour-based features in hue saturation and luminance (HSL) channels which were fed as input to an

MLP classifier that was optimized using genetic algorithms (GAs). The classifier was able to classify the apples into their respective classes: unripe, half-ripe, ripe, and overripe with an accuracy of 97.88 %. A 93.02 % classification accuracy was achieved (Castro et al., 2019) by combining colour space conversion and ML to cape gooseberry fruit classification. They extracted the mean values of RGB images as colour parameters and converted them to the $L^*a^*b^*$ colour space. Then, the colour features were combined with visual features for training on several ML algorithms, of which the SVM emerged as the best model. Steinbrener et al. (2019) captured images of different hard fruits using 16 spectral bands ranging from 470 to 630 nm, which were compressed into three-band images. These were further trained on the GoogleNet model that was pre-trained on RGB image data to form what they referred to as a kernel model. This model achieved accuracies of 96.55, 88.3 %, 100 %, 97.7 %, and 100 % for apples, avocados, bananas, mangoes, and tomatoes, respectively. Cho et al. (2020) argue that extracting the colour values in $L^*a^*b^*$ reveals firmness and translates to the ripeness of avocado fruit. The $L^*a^*b^*$ values were extracted from images of each avocado fruit collected with a smartphone camera and used as target values for their support vector regression (SVR) algorithm. After training the collected images, the SVR algorithm achieved a root mean squared error (RMSE) of 7.54. Worasawate et al. (2022) investigated the ability of ML algorithms to classify mangoes based on fruit appearance in terms of average colour and weight along with electrical data like capacitance. Of all the ML algorithms compared, the feed-forward artificial neural network (ANN) emerged as the best model, with an accuracy of 93.6 % for a two-class (ripe, and unripe) classification experiment and 89.6 % for a three-class (ripe, overripe, and unripe) classification experiment.

In (Worasawate et al., 2022) images of strawberries were acquired at 19 spectral bands from 405 nm to 970 nm and principal features were extracted using PCA. The resulting features were trained on several ML algorithms, and 100 % accuracy was achieved with the SVM classification model. Kangune et al. (2019) trained a CNN classification algorithm on extracted contour and histogram features obtained from RGB, and HSV grapefruit images acquired in the field to classify each fruit as either ripe or unripe. Steinbrener et al. (2019) extended their study to include different soft fruits and achieved ripeness accuracies of 100 % for both strawberries and grapes. Similarly, Gao et al. (2020) captured images of strawberry fruits using spectral bands ranging from 370 nm to 1015 nm. The spectral wavelengths of 530 nm and 528 were identified to clearly define bands of early ripe and ripe strawberries in the field and lab, respectively. Using PCA, they extracted three principal components and combined them as a three-channel image for detecting early ripe and ripe strawberries. These were further processed by training on an AlexNet model, pre-trained on the RGB ImageNet dataset, and achieved 98 % and 100 % for field and lab-collected images, respectively. Similarly, SVM and PCA were combined in Liu et al. (2014) for achieving 100% accuracy in ripeness classification of multispectral strawberry images acquired at wavelengths in the range of 405–970 nm. In (Raj et al., 2022) they extracted spectral signatures related to strawberry fruit ripeness using a narrowband hyperspectral spectroradiometer. Using the spectral reflectance images as input to an SVM classifier, the proposed model was able to discriminate between ripe and unripe strawberries and achieved 98.2 % accuracy.

Other research works such as (Saranya et al., 2021; Miragaia et al., 2021) applied DL algorithms directly to RGB images to estimate the ripeness levels of fruits. In (Saranya et al., 2021), a CNN architecture was used to classify bananas into ripe, partially ripe, ripen, and overripe classes and achieved a 96.14 % accuracy on the test set. Miragaia et al. (2021) captured images of plum fruit in the field, cropped them to focus on the fruit of interest, and trained on the AlexNet model pre-trained on the ImageNet with modifications made to the dense layers. The proposed method in (Miragaia et al., 2021) was able to differentiate between three varieties of plums and classify them based on their levels of ripeness with a 92.83 % accuracy.

Some studies have extracted spectral indices generated by averaging pixels in each segmented fruit region. Others extracted colour-based features, such as mean, standard deviation, and skewness values of intensity, colour moments, and/or histograms for training a classifier. The colour-based features are best established using laboratory-controlled images, given that stability is more likely to be achieved under such structured conditions rather than in field settings. The ripeness estimation models which used images captured in the visible range (Saranya et al., 2021; Miragaia et al., 2021; Raj et al., 2022) achieved good accuracies. Particularly, the high-performance accuracies of Vis-NIR images Gao et al. (2020), have shown they are better at overcoming the rigorous processes of extracting spectral features or visible colour-based features in the form of indices. Even though these lines of thought present a promising direction for fruit ripeness estimation, neither visible range captured images nor colour-based features are suited to the challenges that mature blackberry fruit presents. The traits of ripeness in blackberry fruits at a mature stage are very subtle and cannot be discerned at the visible range. Therefore, it is necessary to explore the analysis of spectral images to advance the research in fruit ripeness assessment. The significance of spectral images compared to spectral indices has been shown by (Hong et al., 2023) to be substantially advantageous in spectral analysis, though for remote sensing downstream tasks such as single/multi-label pixel-wise scene classification, semantic segmentation, and change detection.

This paper proposes a non-destructive approach to in-field blackberry ripeness assessment by utilizing the capability of hyperspectral imaging and deep learning for detecting the traits of ripeness in a mature blackberry fruit under in-field conditions. The former involves the utilization of a line-scanning hyperspectral camera covering wavelengths ranging from 600 nm to 975 nm to experimentally determine the spectral wavelengths that best capture the reflectance characteristics of ripeness in mature blackberry fruit. This is to be used for acquiring images at the determined wavelengths instead of exploring the spectral indices of ripeness. The latter comprises the design of a novel multi-input CNN ensemble classifier that is capable of learning the intrinsic ripeness characteristics of mature blackberry fruit images. The multi-input CNN ensemble model optimises the fully connected layers of the VGG16 architecture pre-trained on the ImageNet dataset and serves as a base for the ensemble learners. The homogeneous ensemble learning with stack generalization ensemble (SGE) framework is explored. To the best of our knowledge, this is the first study to investigate automated ripeness detection of blackberry fruits, and in a field environment. This paper's contributions includes:

- Customizing a stereo camera lens to filter light at wavelengths determined via analysis of acquired on-farm hyperspectral images.
- Design of a novel multi-input CNN_{vgg16} ensemble (MCE) architecture that adapts pre-trained weights to extract features that represent ripeness characteristics of mature blackberry fruits acquired at determined wavelengths. By optimising the fully connected layers of the network, their concatenated features are leverage for building a homogenous ensemble learners' base model. Using a meta-learning algorithm within an SGE framework, the prediction results of each learner are leveraged for predicting the final class of a mature blackberry fruit, as ripe or unripe. The proposed model offers a new insight into fruit ripeness.
- Adoption of a qualitative statistical method, like Pearson correlation to analyse the strength of correlation between the machine sensory and human sensory in fruit ripeness estimation.

The rest of this paper is organized as follows. The proposed methodology is presented in Section 2 and Section 3 is the experimental settings. Section 4 presents the experimental results and discussions while Section 6 concludes the paper.

2. Materials and methods

2.1. Data collection

The experimental dataset consists of blackberry fruits acquired from two commercial blackberry farms located in the UK – one is the Windmill Hill Fruits Ltd in Herefordshire (52.0564° N, 2.7160° W), UK, while the other is the Clockhouse Farm in Kent (41.71° N 73.48° W), UK. In this paper, these farms are identified as Farm-A and Farm-B, respectively. To select the spectral wavelengths that correspond to ripeness levels of blackberry fruit, the categories of raw (0), unripe (1), near ripe (2), ripe (3), and overripe (4) blackberry fruits are sampled and imaged using the IMEC Ltd (Gent, Belgium) line scanning hyperspectral camera which covered wavelengths from 600 nm to 975 nm. The camera operated under spatial resolution of 2048xlength and spectral resolution less than 10 nm. The camera was set-up inside a stainless frame with six 12 Watts tungsten halogen lamps. Since this was a laboratory-driven experiment, the variability in spectral reflectance due to environmental factors (Hong, et al., 2018) remained unknown. It can be argued that if a fruit ripeness assessment model is capable of disentangling ripeness in an image of a blackberry fruit from constituent noise due to spectral variability, then the effect can be negligible. This is based on the fact that though the spectral variability introduces a decrease in reflectance spectra, the spectral wavelength remains the same as is seen in (Hong, et al., 2018). However, steps were taken to ensure only the blackberry fruit reflectance is considered. After imaging, subsequent processing include segmentation to eliminate background noise which is preceded by averaging of the intensities of the segmented berries to generate ripeness reflectance spectrum through a process termed linear mixing. Fig. 1 shows a significant difference in spectral reflectance of blackberry ripeness which ranges from 600 nm to 850 nm with the most significant from 700 nm to 850 nm. The ripeness characteristics of mature blackberry fruits, near ripe (2) and ripe, show close reflectance characteristics over the spectral bands, even at near-infrared bands. However, a striking difference in the ripeness characteristics can be observed at a wavelength of 770 nm. Therefore, this paper explores the near-infrared spectral wavelength that best reflects ripeness characteristics and a visible spectral wavelength to maximize their respective cues for detecting ripeness in mature berries.

To acquire images in the Vis-NIR (bispectral images) wavelengths, an Arducam synchronized stereo 12 MP camera is fitted with filters at spectral wavelengths of 700 nm and 770 nm, respectively. The stereo cameras worked in synchrony, as can be seen in Fig. 2. Fig. 2 shows the on-farm use of the stereo camera, and its resulting image is a horizontally aligned and placed side-by-side to each other. Since the challenge for pickers is discerning between a mature ripe berry and a mature unripe berry or over-ripe berry, only the mature berries were captured. The mature ripe berry implies blackberry fruits that show the appearance and colour of a ripe berry and taste sweet, whereas the mature unripe berry implies blackberry fruits that show the same appearance and colour of a ripe berry but taste sour. In all, a total of 87 and 214 blackberry fruit samples were collected from Farm-A and Farm-B, respectively. An RGB camera is used alongside the stereo camera for performance comparison purposes.

2.2. Multi-input CNN_{vgg16} ensemble learner

A multi-input CNN_{vgg16} ensemble learner (MCE) is proposed for classifying a blackberry fruit as ripe or unripe. This learner combines a multi-input CNN_{vgg16} architecture and ensemble learning.

2.2.1. Multi-input CNN_{vgg16}

The multi-input CNN_{vgg16} architecture proposed in this work is structured on the strength of VGG16 (Simonyan and Zisserman, 2014) pre-trained on ImageNet. An illustration of the architecture can be observed in Fig. 3. The decision to use a pre-trained VGG16 model comes

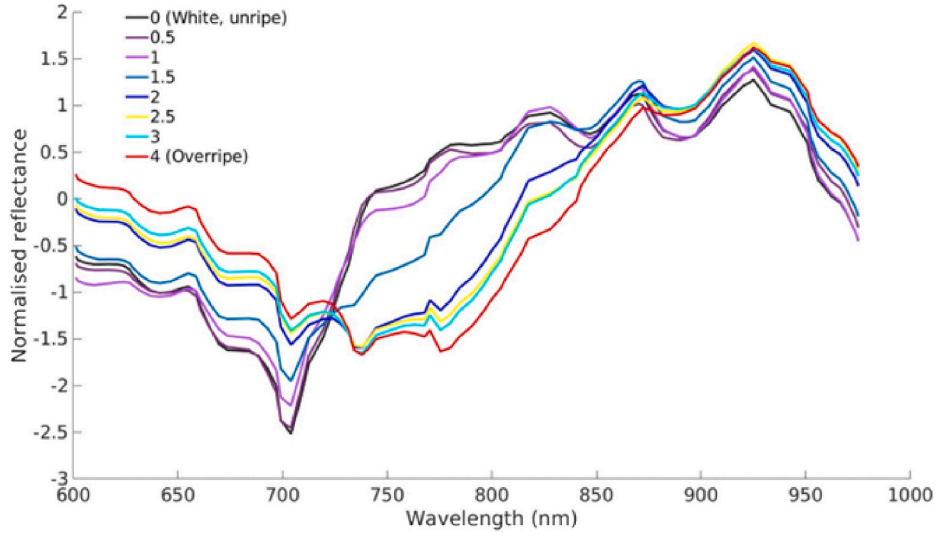


Fig. 1. Normalised reflectance spectra at different blackberry ripeness levels.

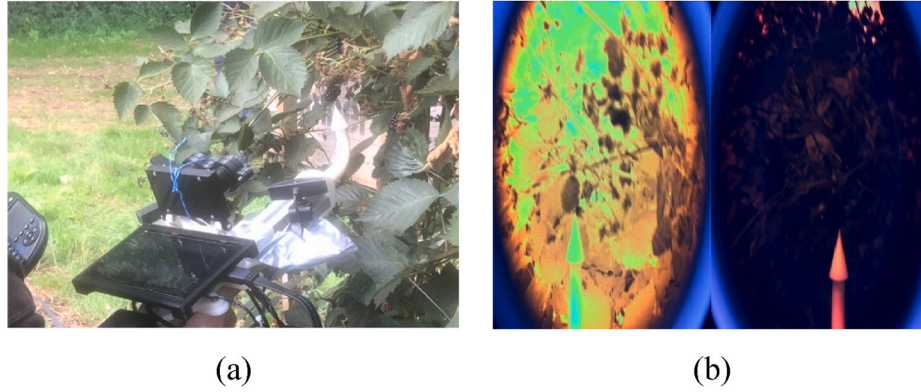


Fig. 2. On-farm blackberry imaging using multiple-camera mount with a pointer to aid fruit identification. a) stereo camera and RGB camera, b) a horizontally aligned image which is a sample of the stereo camera imaging.

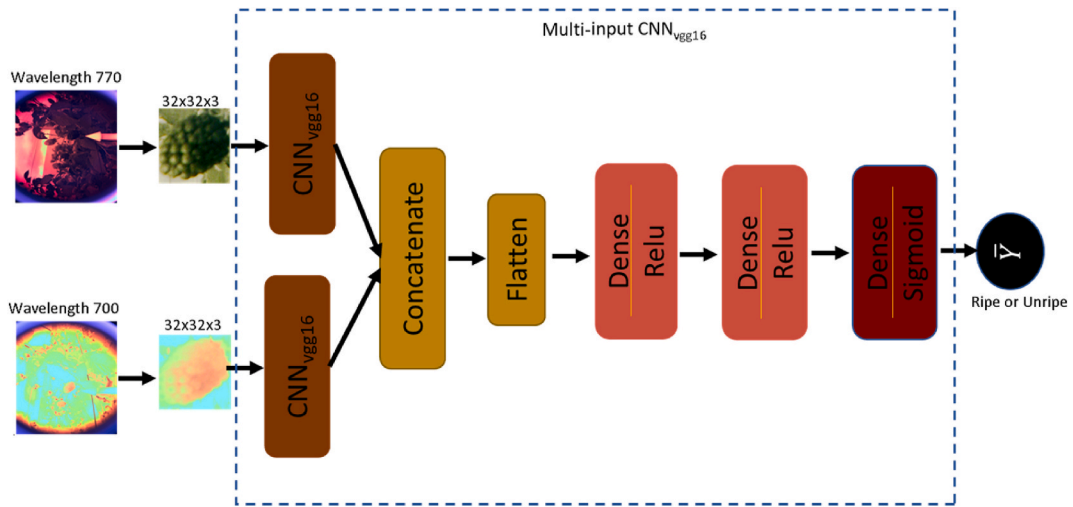


Fig. 3. The proposed multi-input CNN_{vgg16} classifier for detecting ripeness in blackberry fruits. The input to the network is a pseudo-coloured $32 \times 32 \times 3$ bispectral blackberry sample.

from its performance in comparison to other pre-trained models, such as DenseNet (Huang et al., 2017), Xception (Chollet, 2017), InceptionV3 (Szegedy et al., 2016), and MobileNet (Howard et al., 2017). They are

chosen based on their complexity and depth. As observed through the experimental results in Table 1 and Fig. 5, the lesser the complexity and depth, the better the transferability from the source task to the target

Table 1
Evaluation of different CNN models for blackberry ripeness detection.

Model	Depth	Image Size	FC Layer	Precision	Recall	F1-Score	Test Accuracy
DenseNet	242	$32 \times 32 \times 3$	$N^{512,256}$	0.872	0.735	0.772	0.878
InceptionV3	189	$75 \times 75 \times 3$	$N^{256,64}$	0.808	0.829	0.797	0.829
Xception	81	$71 \times 71 \times 3$	$N^{512,256}$	0.777	0.683	0.712	0.683
MobileNet	55	$75 \times 75 \times 3$	$N^{512,256}$	0.873	0.878	0.875	0.878
Vgg16	16	$32 \times 32 \times 3$	$N^{1024,1024}$	0.898	0.902	0.897	0.902

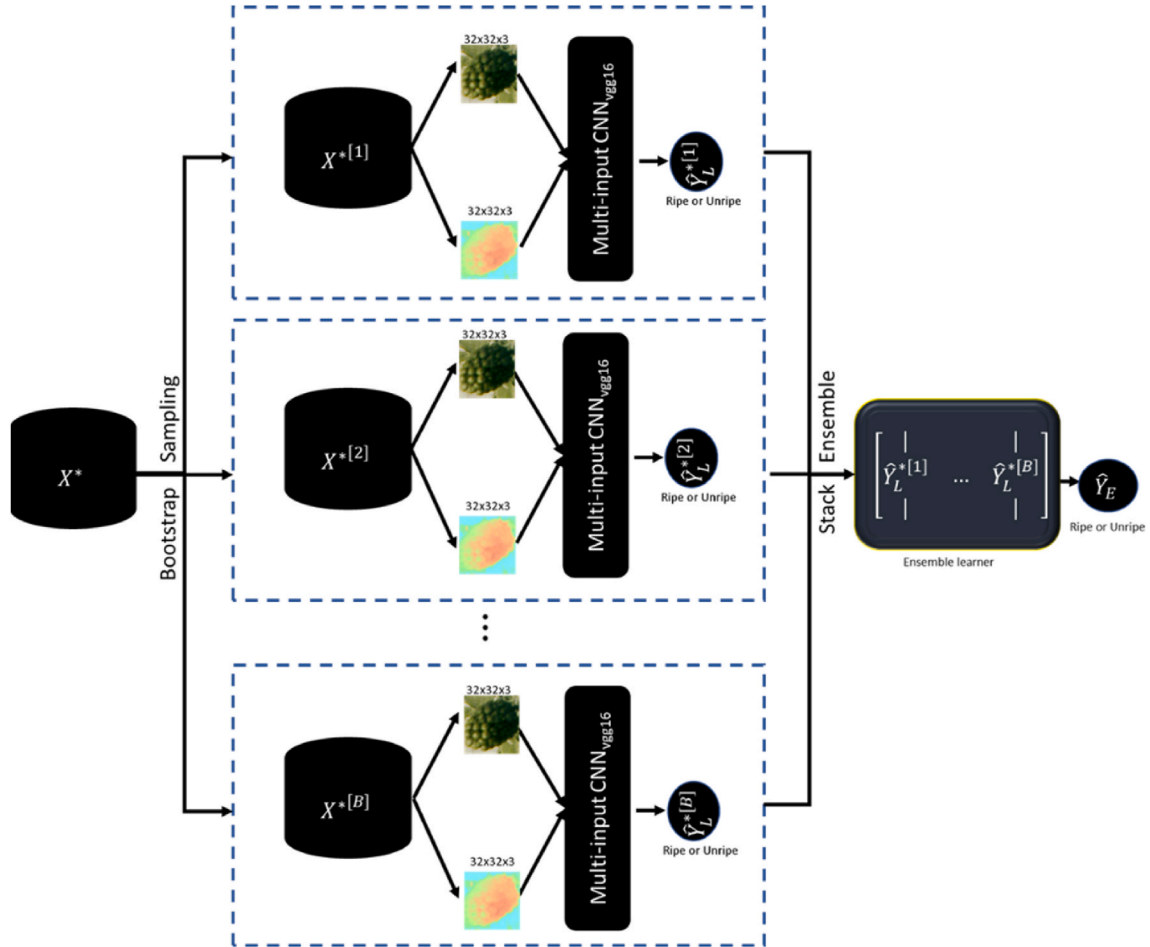


Fig. 4. Homogenous ensemble of learners with stacking architecture. Using the blackberry fruit dataset, a bootstrap subset is created through bootstrap re-sampling with the multi-input CNN_{vgg16} as base learner.

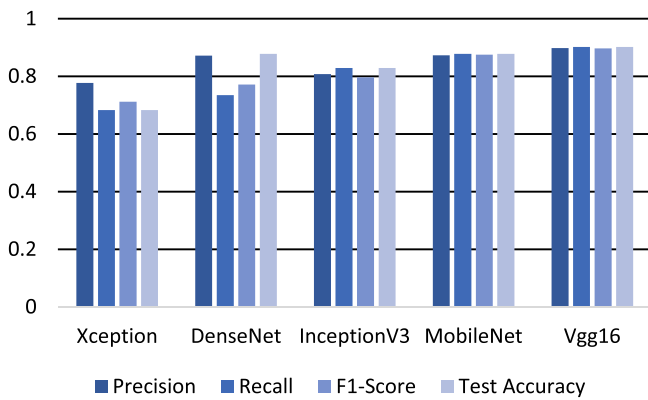


Fig. 5. Comparison of pre-trained CNN models performance on blackberry ripeness detection. The VGG16 emerged the best.

task. Though the experiment was constrained because it is task-specific, the grounds of the empirical analysis suggest the reasons might be because the target task is characterized by locally situated and less complex objects than the source task, and 2) the target task is of small sample size data. A summary of modifications made to the networks is presented in Table 1.

As conventionally done in transfer learning, the learned weights generated at the convolutional layers from training VGG16 on the ImageNet dataset is frozen. Then through weight update at the fully connected (FC) layers, the network is able to learn the discriminative features of a mature ripe/unripe blackberry fruit. The VGG16 configuration (Simonyan and Zisserman, 2014) changes with 1) the type of input to the network which in this case is the stereo (bispectral) images acquired with camera filters set to spectral wavelengths of 700 nm and 770 nm. It has been shown that knowledge transfer from a domain of a different task and image format to another is feasible in transfer learning (Olishah and Smith, 2019); 2) the size of the input to the network is critical to the design. A $32 \times 32 \times 3$ image size is considered to minimize

computational cost. The stereo camera introduced a slight positional difference, known as stereo disparity, between the berries and combining the images in the spatial domain distorts the berries. The registration techniques are ineffective in this case because of the varying degrees of intensity between the stereo images. Therefore, concatenation of features in the feature space is preferred and it is achieved by a downsizing the image; 3) the number of neurons of the FC layers. The number of neurons is dynamically determined through stratified k-fold cross-validation to ensure the number of neurons in the FC layer is capable of learning the ripeness cues of blackberry fruits, and 4) the best hyperparameters which are likewise determined through stratified k-fold cross-validation for the best configuration of the network. To compile the multi-input CNN_{vgg16} training algorithm, the rectified linear unit (ReLU) activation function (Agarap, 2018), is preferred because it overcomes the vanishing gradient problem. The classification layer uses the sigmoid activation function to output values in the range [0,1], where values ≥ 0.5 indicate a mature unripe berry and values < 0.5 signify a mature ripe berry. During training, early stopping is combined with model checkpoint to overcome overfitting. Overfitting results from the inability of the network to learn the relationship of significant within-class features, but rather learns the noise in the data. Transfer learning can address the overfitting problem, but the transferability of features decreases with dissimilarity between source task and target task (Yosinski et al., 2014).

2.2.2. Ensemble learning

With the best-performing hyperparameters, the multi-input CNN_{vgg16} of about 600 K trainable parameters serves as the base model for ensemble learning and is henceforth named MCE. This paper creates homogeneous ensemble learners, that is, learners that share the same learning architecture as depicted in Fig. 4. The concept of ensemble learning was first introduced in 1965 (Nilsson, 1965) and has gained popularity in the classification task, see for instance (Liu et al., 2017; Shahhosseini et al., 2021).

Given blackberry dataset, \vec{X} , a bootstrap subset $(X^{*[1]}, X^{*[2]}, \dots, X^{*[B]})$ is created through bootstrap resampling (Efron, 1992). This method involves resampling (with replacement) from a sample dataset to create the bootstrap subsets. Since the base learners were not generated by cross-validation, the minority class of the blackberry dataset is oversampled using the process of random oversampling (Menardi and Torrelli, 2014). Each bootstrap subset is trained using the base learner, multi-input CNN_{vgg16}, and is mathematically expressed as:

$$\hat{Y}_L^{*[i]} = \varphi \left(\sum_{j,k=1} w_{j,k} \left(x_{j,k}^{*[i]} \right) + b \right) \quad i = \{1, 2, \dots, B\}, j = \{1, 2, \dots, n\}, k = \{1, 2, \dots, n\} \quad (1)$$

where $x_{j,k}$ is the j^{th} and k^{th} vectorized bi-spectra samples from a bootstrap subset, $w_{j,k}$ represents the weights, b is the bias and φ is the activation function. Usually, the outcome, \hat{Y} , of the base learners, L , for i^{th} bootstrap subset, should be combined into a single prediction result, hence it is important to consider the integration technique applied to the base learner's prediction.

Several base learners' integration techniques exist, including bagging, boosting, and stacking (Sewell, 2008). The bagging method assigns equal weight to all base learners' predictions, while the boosting method adjusts the weight of the training data of a prior base learner to improve on the prediction of a subsequent base learner. However, if a preceding base learner poorly performs, the performance of subsequent base learners cannot be guaranteed. Even though the multi-input CNN_{vgg16} architecture is optimized, which makes it appropriate for bootstrapping, the stacked generalization ensemble (SGE) framework, proposed by Wolpert in 1992 (Wolpert, 1992), is adopted. SGE combines the predictive accuracy of base learners to create a more powerful predictive ensemble learner. The SGE framework works by stacking the

prediction results of base learners as feature vectors and uses a meta-learning algorithm to make a single prediction that is expected to outperform any of the base learner's predictions. The meta-learning algorithm used is the logistic regression algorithm, and the process is mathematically expressed as:

$$\hat{Y}_E = \frac{1}{1 + e^{-(\beta_1 \hat{Y}_L^{*[i]} + \beta_0)}} \quad (2)$$

where $\hat{Y}_L^{*[i]}$ is a column feature vector with each column representing a base learner's, L prediction, β_1 is the regression coefficient of the feature vectors estimated using a method generally known as maximum likelihood estimation, β_0 is the intercept of the feature vectors, and \hat{Y} is the outcome of the ensemble learner, E .

3. Experimental settings

3.1. Dataset

The blackberry fruit dataset comprises of blackberries assessed subjectively through fruit quality assessment, with and without an expert assessors. The expert assessors involvement at some point in the data collection is to analyse the performance of machine sensory, MCE, in comparison to the human sensory. Therefore, the following question is posed: Can the proposed blackberry ripeness detection algorithm differentiate between a mature ripe and unripe berry with comparable performance to laboratory-level fruit quality assessment based on human sensory evaluation? If the proposed blackberry ripeness detection algorithm performs comparably in the field or outperforms laboratory-level fruit quality assessment, then blackberry growers will be able to overcome the need to transport blackberries to the laboratory and can minimize the number of mature unripe blackberries making their way to the market.

In all, the blackberry fruit images in RGB and bispectral formats from Farm-A and Farm-B used for the experiments are labelled as follows, FA_{RGB}, FB_{RGB}, FA_{bispec}, and FB_{bispec}, and comprises of a total of 87, 97, 82, and 119 samples, respectively. Their combination is labelled as FA + B_{RGB} and FA + B_{bispec} with a total of 187 (148 ripe and 39 unripe) and 201 samples (164 ripe and 37 unripe), respectively, and split into 164 and 37 for the class of mature unripe berries. The bispectral and RGB sets are cropped to the blackberry of interest and resized. To improve the contrast of berries captured at a spectral wavelength of 770, the histogram equalization technique is applied. It should be noted that the assessment of the blackberry samples from Farm-A are without fruit quality expert involvement, while blackberry samples from Farm-B required assessment by fruit quality experts. The reason the total number of samples used for the experiment differs from the total number collected from both farms is due to the following:

- i. A visible unripe appearance – images of a berry that show visible unripe attributes such as a presence of green or red colour are disregarded because such berries are not yet mature and therefore are outside the scope of the project.
- ii. Motion blur – images were acquired during the harvesting season as pickers were harvesting. This resulted to some images being greatly affected by motion blur due to camera movement. For this reason, these images were discarded.
- iii. No reference berry – these are samples the pointer showed no reference berry.

3.2. Classifier evaluation and optimization

The experimental scenarios for evaluating the proposed algorithm are comprehensively presented as follows: 1) classification experiments on FA_{RGB}, FB_{RGB}, FA_{bispec}, and FB_{bispec}, FA + B_{RGB}, and FA + B_{bispec} using our multi-input CNN_{vgg16} classifier without optimization, 2)

Classification experiments on FA + B_{bispec} using our multi-input CNN_{vgg16} classifier with optimization, 3) Classification experiments on FA + B_{bispec} using the optimized multi-input CNN_{vgg16} with ensemble learning, and 4) proposed machine learning-based ripeness detector versus laboratory level fruit quality assessment using FB_{bispec}. For each scenario, the data is split into training, validation, and testing sets in the ratio of 60:20:20. The performance metrics are weighted precision, weighted recall, weighted f1-score, and accuracy. Accuracy computes the ratio of correct prediction of mature ripe and unripe blackberries to the total number of mature blackberries. Weighted precision finds the ratio of mature ripe blackberries correctly predicted by computing the weighted percentage of true positives, (TP), for each class to the number of mature blackberries predicted, (TP), and false negative (FP), as mature ripe ones. Weighted recall finds the ratio of the number of mature ripe blackberries correctly predicted by computing the weighted percentage of true positives, (TP), for each class to the number of mature ripe blackberries comprising (TP) and false negatives (FN). The weighted f1-score is computed as the weighted average of weighted precision and weighted recall. These are mathematically expressed as follows.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

$$Precision_w = \frac{TP_{weighted}}{TP + FP} \quad (4)$$

$$Recall_w = \frac{TP_{weighted}}{TP + FN} \quad (5)$$

$$F1 - score_w = 2 * \left[\frac{(Recall_{weighted} * Precision_{weighted})}{Recall_{weighted} + Precision_{weighted}} \right] \quad (6)$$

The stratified random sampling cross-validation method is adopted for splitting of FA_{RGB}, FB_{RGB}, FA_{bispec}, and FB_{bispec}, and FA + B_{RGB}, FA + B_{bispec} data into training, validation, and test sets. This method is for addressing the class imbalance between mature ripe and unripe blackberry samples and to ensure that a balanced class ratio is maintained for each split. The optimization experiment uses the grid search algorithm for fine-tuning hyperparameters of the multi-input CNN_{vgg16} classifier with the stratified *k*-fold cross-validation method. Given that the proposed classifier is of a multi-input structure, the grid search with a stratified *k*-fold cross-validation approach is implemented differently. For each experiment with a hyperparameter, other hyperparameters remain unchanged, then a discrete grid of hyperparameter values is set such that the stratified *k*-fold cross-validation method is implemented for each value of the grid. Using the stratified *k*-fold cross-validation method, the data is divided into *k* subsets to ensure that the class ratio remained the same in each fold - a fold for testing and the remaining *k*-1 (*k* = 10) folds for training and validation. The performance of a hyperparameter is evaluated by averaging the results obtained in each *k*-1 experiment, then the process is repeated for a different value of the hyperparameter in the grid. In this paper, the fine-tuned hyperparameters are the number of neurons in the FC layer, optimization functions, batch size, and epoch. The descriptions and value settings of these hyperparameters, as used in the experimental scenario (1) for the combined dataset, are presented in Table 2, while the hyperparameters for the experiment scenario (2) are provided alongside their results in Table 4.

4. Experimental results and discussion

Experiments were carried out in the order of the experimental scenarios and therefore results, and discussions, are reported accordingly.

Table 2

The multi-input CNN_{vgg16} classifier hyperparameters.

Hyperparameter	Description	Initial Values
FC Layer	The fully connected (FC) layer defines the number of neurons connected to each input from the last convolutional layer of the pre-trained VGG16 network for the weights of the target task to be learned.	N ^{1024,1024}
Optimization Function	An algorithm that minimizes the loss function of the network during training.	Adam
Batch Size	The number of samples utilized in one iteration.	10
Epoch	Defines the number of passes made through the entire training dataset during training.	20

Table 3

The ripeness detection performance of the multi-input CNN_{vgg16} classifier on different farm datasets.

Data Set	Precision	Recall	F1-Score	Train Accuracy	Test Accuracy
FA _{rgb}	0.556	0.556	0.556	0.783	0.556
FB _{rgb}	0.700	0.700	0.700	0.974	0.699
FA _{bispec}	0.882	0.824	0.827	0.985	0.824
FB _{bispec}	0.889	0.893	0.889	0.938	0.893
FA + B _{rgb}	0.623	0.789	0.696	0.792	0.789
FA + B _{bispec}	0.898	0.902	0.897	0.963	0.902

4.1. Multi-input CNN_{vgg16} classifier without optimization

The results of applying the multi-input CNN_{vgg16} classifier on the FA_{RGB}, FB_{RGB}, FA_{bispec}, FB_{bispec}, FA + B_{RGB}, and FA + B_{bispec}, respectively, are presented in Table 3 and Fig. 6. It can be observed in Table 3 and Fig. 6 that the multi-input CNN_{vgg16} classifier which takes as input the bispectral images of a blackberry fruit acquired at two spectral wavelengths, 700 and 770 nm, can detect the ripeness of mature blackberry fruits than the RGB images from Farm-A and Farm-B. The average level of performance of the RGB-based model for both individual and combined Farm-A and Farm-B datasets is not surprising because mature ripe and unripe blackberry fruits are not visually discernible from each other.

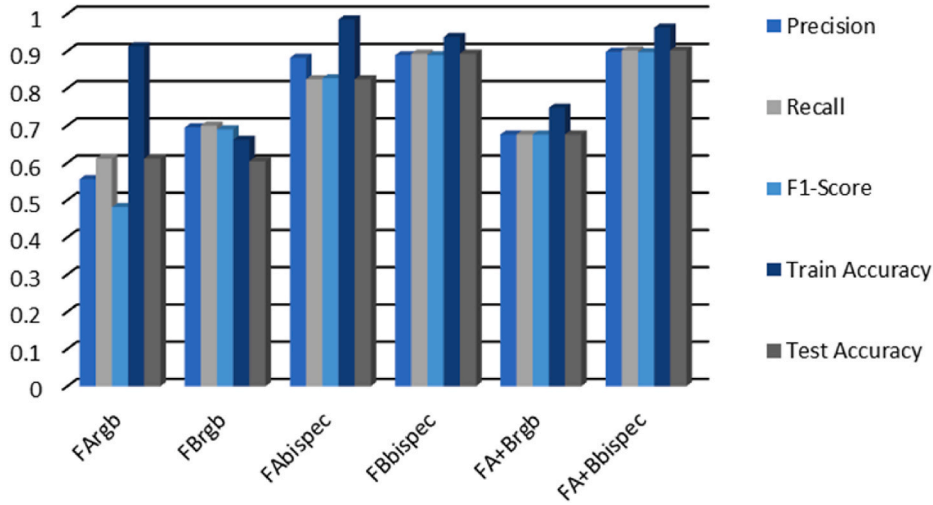
The multi-input CNN_{vgg16} model can be observed in Table 3 to overfit to FA_{RGB} and FB_{RGB} training sets, but it is addressed with more examples of ripe and unripe class when the combined FA + B_{RGB} data set is used. However, accuracy remained a little above average and skewed toward the ripe class. On the other hand, a stable performance is observed with FA_{bispec} and FB_{bispec} data sets. The 6.9 % increase in accuracy achieved with the FB_{bispec} set might be because the samples from Farm-B were subjectively assessed by expert fruit assessors and therefore were more correctly labelled.

With the bispectral combined dataset, FA + B_{bispec}, the multi-input CNN_{vgg16} achieved 90.2 % accuracy. However, precision and recall present interesting accuracies, which are 90.2 % and 89.8 %, respectively. Given that the mature unripe berries constitute about 20 % of the set and the remaining 80 % for the mature ripe berries, both metrics are useful for evaluating the model's performance. The precision gives more insight into the performance of the model on the unripe class, while the recall on the ripe class. A 90.2 % recall shows the model correctly classified the ripe class, and a precision value of 89 % shows the model performed equally well on the unripe class. It should be noted that the true class was assigned to the unripe blackberry fruits and the ripe blackberry fruits, the negative class. This ordering was automatically defined by the order of the sample folders, but does not in any way affect the result of the model. Overall, it can be concluded that the spectral filters of wavelengths, 700 nm, and 770 nm, are useful for uncovering the hidden ripeness features of mature berries.

Table 4

The Multi-input CNNvgg16 hyperparameter optimization evaluation. The best hyperparameter feeds into the next experiment.

Hyperparameter	Value	Precision	Recall	F1-Score	Train Accuracy	Test Accuracy
FC Layer	$N^{256,256}$	0.910 ± 0.048	0.905 ± 0.051	0.893 ± 0.061	0.959 ± 0.039	0.905 ± 0.048
	$N^{512,512}$	0.913 ± 0.044	0.900 ± 0.056	0.888 ± 0.064	0.961 ± 0.032	0.900 ± 0.053
	$N^{1024,1024}$	0.921 ± 0.041	0.917 ± 0.038	0.911 ± 0.043	0.968 ± 0.029	0.915 ± 0.049
	$N^{2048,2048}$	0.920 ± 0.035	0.909 ± 0.042	0.899 ± 0.056	0.964 ± 0.024	0.895 ± 0.048
	$N^{256,256}$	0.883 ± 0.099	0.898 ± 0.060	0.882 ± 0.079	0.961 ± 0.048	0.898 ± 0.058
	$N^{512,256}$	0.923 ± 0.048	0.922 ± 0.043	0.914 ± 0.048	0.973 ± 0.017	0.917 ± 0.037
	$N^{1024,256}$	0.887 ± 0.096	0.900 ± 0.052	0.887 ± 0.073	0.963 ± 0.049	0.888 ± 0.043
Optimization Function	$N^{2048,256}$	0.876 ± 0.087	0.887 ± 0.045	0.873 ± 0.064	0.961 ± 0.048	0.900 ± 0.039
	Adam	0.915 ± 0.056	0.913 ± 0.054	0.910 ± 0.034	0.970 ± 0.029	0.915 ± 0.060
	SGD	0.872 ± 0.084	0.876 ± 0.044	0.845 ± 0.069	0.927 ± 0.043	0.876 ± 0.041
Batch Size	AdaGrad	0.799 ± 0.105	0.826 ± 0.017	0.764 ± 0.037	0.861 ± 0.022	0.827 ± 0.017
	4	0.909 ± 0.049	0.907 ± 0.044	0.899 ± 0.049	0.972 ± 0.018	0.907 ± 0.042
	6	0.903 ± 0.049	0.902 ± 0.047	0.893 ± 0.056	0.963 ± 0.033	0.902 ± 0.045
	8	0.910 ± 0.052	0.907 ± 0.049	0.897 ± 0.058	0.966 ± 0.038	0.807 ± 0.047
	10	0.902 ± 0.058	0.897 ± 0.057	0.890 ± 0.067	0.966 ± 0.028	0.897 ± 0.053
	12	0.909 ± 0.054	0.905 ± 0.051	0.893 ± 0.059	0.963 ± 0.031	0.910 ± 0.048
	14	0.903 ± 0.054	0.900 ± 0.053	0.888 ± 0.063	0.965 ± 0.032	0.900 ± 0.051
Epoch	16	0.918 ± 0.056	0.917 ± 0.054	0.911 ± 0.057	0.975 ± 0.018	0.917 ± 0.051
	20	0.909 ± 0.053	0.907 ± 0.047	0.903 ± 0.050	0.971 ± 0.032	0.902 ± 0.048
	40	0.957 ± 0.026	0.954 ± 0.029	0.949 ± 0.036	0.980 ± 0.016	0.919 ± 0.046
	60	0.898 ± 0.039	0.893 ± 0.035	0.879 ± 0.043	0.961 ± 0.029	0.893 ± 0.033
	80	0.955 ± 0.032	0.951 ± 0.035	0.945 ± 0.042	0.948 ± 0.058	0.885 ± 0.053
	100	0.923 ± 0.054	0.919 ± 0.048	0.914 ± 0.078	0.964 ± 0.028	0.907 ± 0.032

**Fig. 6.** The Multi-input CNN_{vgg16} performance on different farms, Farm-A and Farm-B, datasets.

4.2. Multi-input CNN_{vgg16} classifier with optimization

The multi-input CNN_{vgg16} hyperparameters are optimized through grid search and stratified k-fold cross-validation. The hyperparameter, its corresponding values and performance across different metrics, are presented in Table 4 and discussed in subsequent paragraphs. It should be noted that the best value of each hyperparameter, based on the performance metrics - precision, recall, and f1-score, feeds into the multi-input CNN_{vgg16} for computing the next hyperparameter.

To determine the best hyperparameters for the FC layers of the multi-input CNN_{vgg16} classifier, the experiments included fine-tuning the FC layers to, 1) have the same number of neurons at each FC layer, and 2) to vary the number of neurons of the second FC layer while keeping constant the number of neuron in the second FC layer. It is interesting to observe from the experiments that expanding the feature space at the FC layer to twice its input size increased the proposed model's discriminative capability. The best results were recorded across the performance metrics of train accuracy, test accuracy, precision, recall, and F1-score as highlighted in Table 4. The obvious logical explanation might be twice-in-size increases the significance of relevant features. However, with

varying-size FC layer, that is, the first FC layer equalled the size of the input to the layer, and the second FC layer is half its size, higher results is observed across all the performance metrics. By comparison, the precision and test accuracies of the twice-in-size FC layer are 92.1 % and 91.5 %, and the varying-size FC layer are 92.3 % and 91.7 %, respectively. The varying-size FC layer feeds into the multi-input CNN_{vgg16} architecture for obtaining the next best hyperparameter because it performed better by 0.2%.

To determine the best optimization function for the multi-input CNN_{vgg16}, experiments included comparing several optimization functions. They are the Adam optimizer (Kingma and Ba, 2014), stochastic gradient descent SGD (Sutskever et al., 2013), and adaptive gradient AdaGrad (Duchi et al., 2011) optimization functions with their learning rate set to 0.01, 0.01, and 0.001, respectively. Adam being a second-order optimizer uses two exponential decay rates which are set to 0.9 and 0.999 for the first and second moments, respectively. The Adam optimizer performed better in comparison to the SGD and AdaGrad optimizers with a 91.50 % precision. The AdaGrad showed the least performance because it is particularly suited to sparse data.

Using the Adam optimizer, the values of 16 and 40 for batch size and

epoch, respectively, were determined to be the values to best update the weights of the network towards a global minimum. Across the batch size and epoch experiments, the highest scores of 91.80 % and 95.70 %, and 91.70 % and 91.90 %, were observed for precision and test accuracy. .

4.3. Ensemble learning with optimized multi-input CNN_{vgg16}

The MCE model performance can be observed in Fig. 7, Fig. 8 and Table 5. The MCE model achieved a 95.10 % accuracy on the test set, as can be observed in Table 5 which is a 2.8 % increase in performance to the optimized multi-input CNN_{vgg16} model. Using area under the curve (AUC) and precision-recall curve as shown in Fig. 7, the performance of MCE under class imbalance further confirms that the stratified random sampling cross-validation used for building the base model addresses the class imbalance problem. Further, experiment is extended to investigate the impact of unforeseen on-farm environmental factors on MCE model. This is motivated by 1) the varying view angles of the blackberry fruit during harvesting, 2) change in fruit picker viewpoint relative to the fruit position, and 3) changes in environment lighting due to weather conditions (cloudy, sunny, or rainy) or time of day. These factors were modelled using augmentation techniques such as rotation, zoom, and brightness and are significant towards validating the model's capability to withstand unforeseen environmental factors. With a 10° rotation, 0.2 to 1.0 zoom values, and 0.2 to 0.1 brightness values, MCE achieved a performance accuracy of 90.2 % as can be observed in Fig. 8. Even though the complexity of the environmental factors decreased the accuracy of the model, it is still above 90 %.

Further evaluation includes visualising MCE performance using the confusion matrix. Likewise, the model's performance in the presence of environmental factors, as investigated using the augmentation techniques, is visualised using the confusion matrix. The MCE without augmentation is presented in Fig. 9 (a) and with augmentation in Fig. 9 (b-c). The confusion matrix reveals the impact the environmental factors have on the proposed model's performance.

4.4. Human ripeness detection vs machine ripeness detection

Humans have evolved very sensitive and specialized sensory organs, which give them the ability to test fruit quality through a process called sensory evaluation. This innate ability can be further trained, and detailed sensory evaluation can be carried out by expert panels who are

adept at using sensory attributes such as odour, flavour, taste, and texture/appearance to describe and quantify fruit quality attributes. Typically, most blackberry fruit growers would subject some of the harvested berries in any given pick throughout the season to some degree of organoleptic evaluation to ensure that the berries meet consumer requirements. Additionally, Brix is usually measured internally and by or at the request of retailers. In this paper, expert sensory evaluations assessed berries sampled from Farm-B. The evaluation involved a trained evaluator who measured berry fruit quality using sensory attributes and a scoring system as follows: shininess (1 = dull, 5 = shiny), uniformity of colour (1 = uneven, 5 = even), firmness (1 = soft, 5 = firm), skin strength (berry skin broken after manipulation = Y/N), flavor (1 = aromatic, 5 = poor), sweetness (1 = sweet, 5 = sour), texture (1 = melting, 5 = crunchy), and human colour ripeness (0 = green, 1 = red, 2 = black but underripe, 3 = fully ripe, 4 = overripe). Based on these factors, the human ripeness decisions are recorded as *human ripeness* in Table 6.

The machine does not have all the sensory organs as the human, but it tries to compensate for it by maximizing what resembles the visual senses using MCE. This is to enable it to learn characteristics of the blackberry fruit such as points, lines, edges, textures, shapes, and a combination of other factors that correlates to ripeness. To be able to compare machine sensory to the human, the confidence score obtained through evaluating the MCE is used because it measures how confident the model is in its prediction of a given class. There are two classes, ripe and unripe, which are represented as 0 and 1. Each blackberry fruit is labelled under *target* according to *human ripeness*, as given in Table 6.

To qualitatively analyse the MCE performance, the Pearson correlation matrix is observed for MCE confidence against the human sensory variables. As can be observed in Fig. 10, it is evident that the machine sensory is highly and positively correlated to the human sensory evaluation over blackberry fruit texture. A detailed analysis, using Table 6, shows that the texture scores of berries vary with ripeness levels. For example, the machine reached a 99 % confidence score for an unripe berry with a texture score of 5 and a ripeness level of 1.5. Additionally, the human classified one of two unripe berries with texture scores of 4, to be ripe and the other unripe. However, the machine considered both berries to be of different unripe degrees with confidence scores of 78.48 %, and 96.85 %. Machines are designed to complement human efforts in automation and outperform humans in a task requiring efficiency. The MCE model demonstrated this capability in two ways. Berries 3D9262

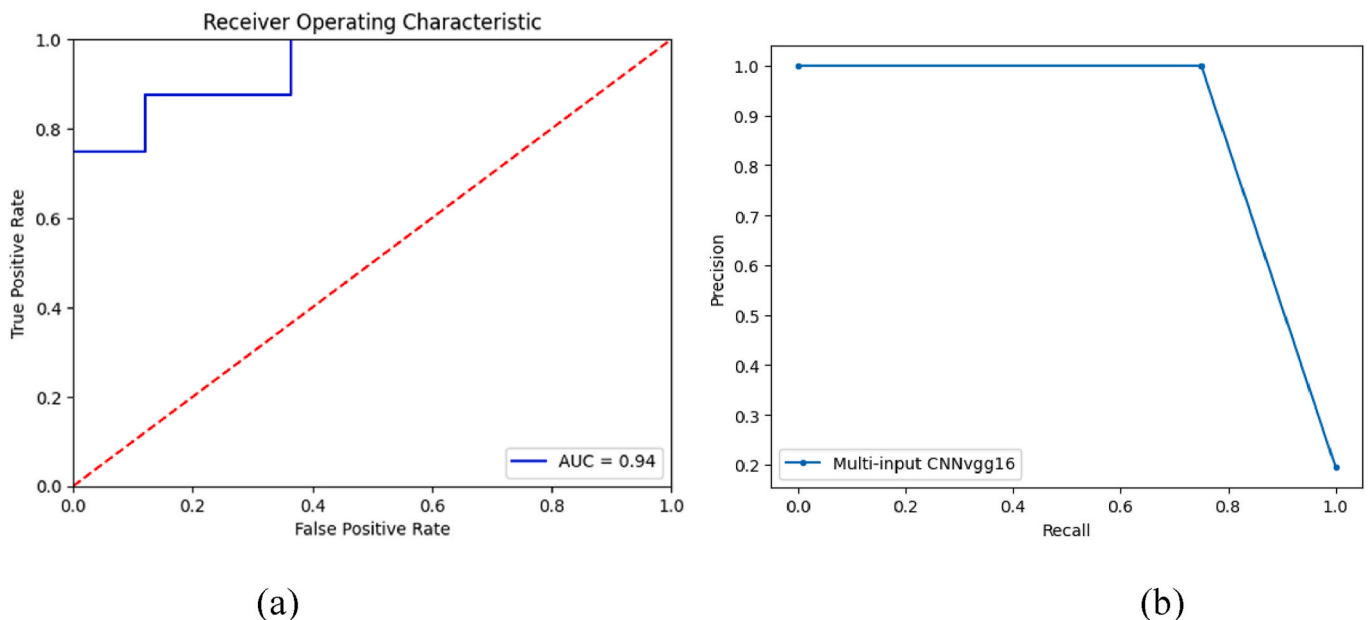


Fig. 7. Evaluation of MCE performance under class imbalance using AUC (a) and precision-recall curve (b).

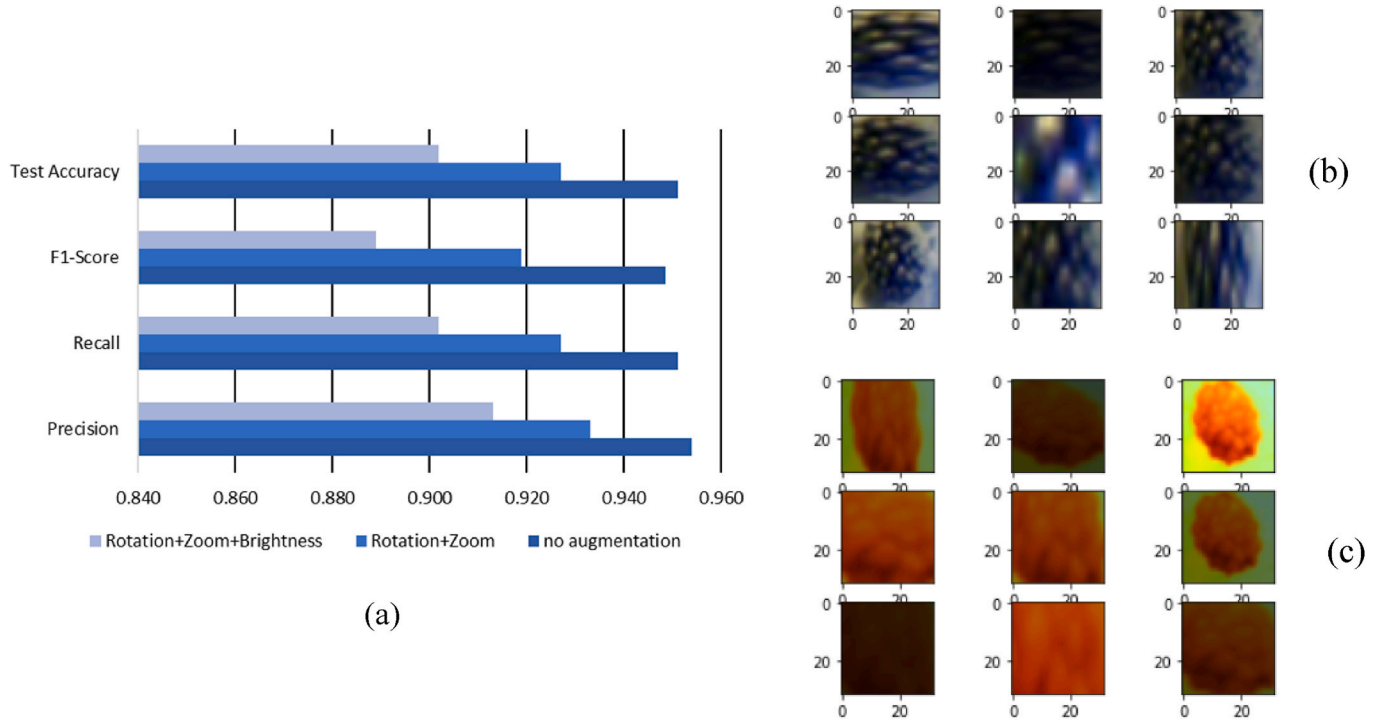


Fig. 8. The MCE performance without and with augmentation. The rotation, zoom and brightness transformations test model capability to withstand unforeseen challenges.

Table 5

Experimental results of the multi-input CNN_{vgg16} ensemble learner with on-farm environmental factors.

State	Precision	Recall	F1-Score	Test Accuracy
No augmentation	0.954	0.951	0.948	0.951
Rotation + Zoom	0.933	0.927	0.919	0.927
Rotation + Zoom + Brightness	0.913	0.902	0.889	0.902

and 976122 were labelled unripe based on their human ripeness scores, however, the machine considered them to be near-ripe by confidence scores of 2.49 % and 1.09 %, respectively. Another scenario where the human scored a berry, 7DE558, as ripe, the machine classified it to be unripe with a 96.85 % confidence score.

5. Conclusion

The traits of ripeness of mature blackberry (*Rubus fruticosus*) fruits are very subtle and renders the task of visual assessment of mature blackberry difficult for growers and pickers. Therefore, this paper proposed a novel multi-input CNN designed from a VGG16 model pre-trained on the ImageNet dataset, and the fully connected layers were optimized for learning and served as the base model for ensemble learning. The ensemble learning comprised homogeneous ensemble learners and their predictions were ensemble using the SGE framework for meta-learning in order to classify a blackberry fruit image as ripe or unripe. The input to the classifier was the blackberry fruits acquired from a stereo sensor with Vis-NIR spectral filters at wavelengths of 700 nm and 770 nm. This paper presented the first research to investigate blackberry fruit ripeness, and the novel multi-input CNN ensemble (MCE) classifier proposed in this paper extends the possibilities of assessing fruit ripeness using hyperspectral images instead of their indices. Through experiments, the proposed model achieved 95.1 %

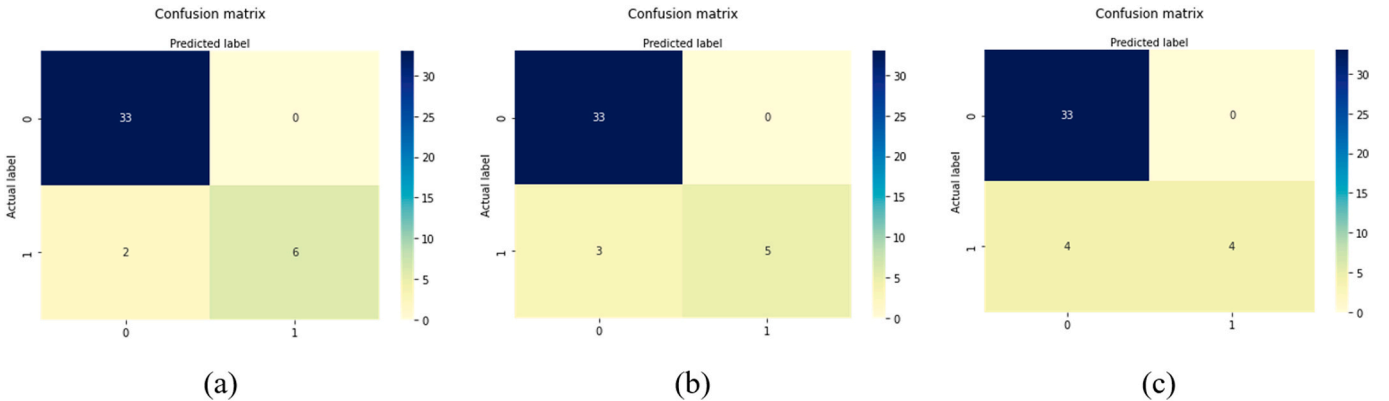
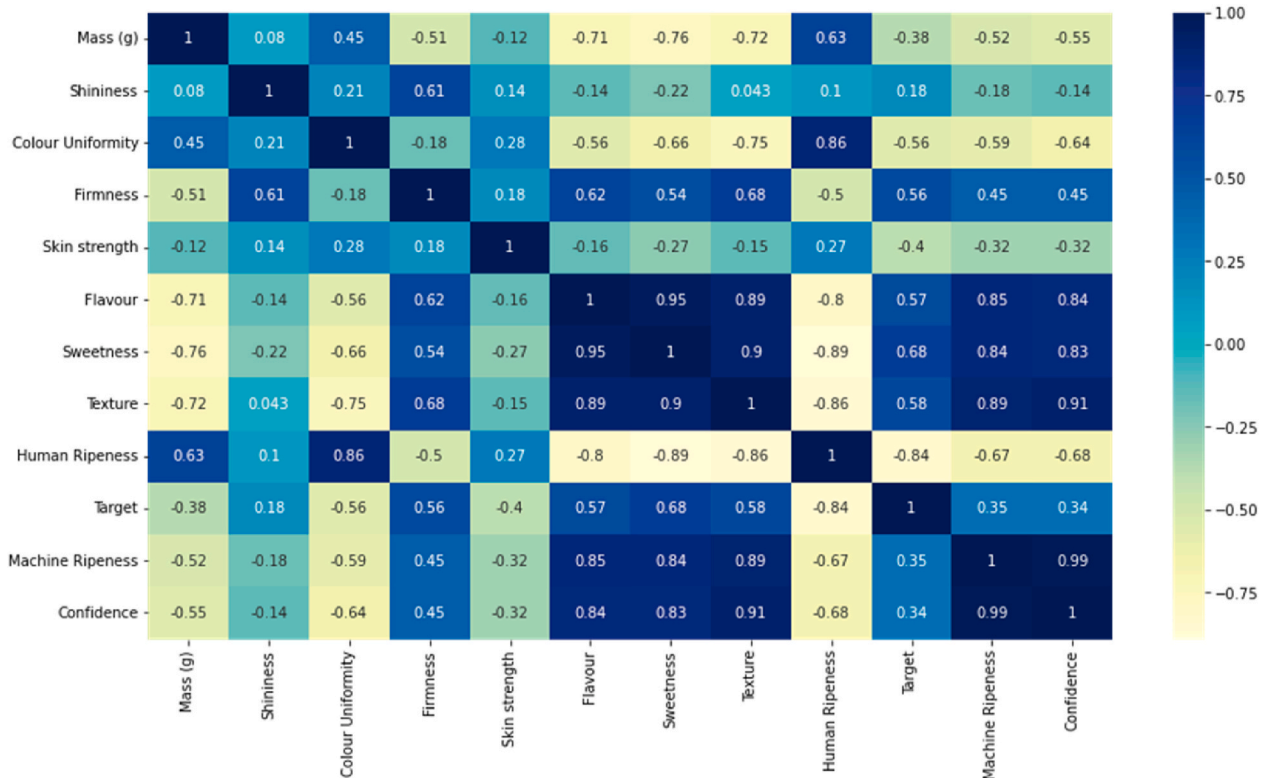


Fig. 9. The confusion matrix of MCE using Farm A + Bbispec dataset. (a) no augmentation (b) with rotation and zoom augmentation, and (c) with rotation, zoom, and brightness augmentation.

Table 6

Comparison of human sensory and machine sensory for blackberry fruit ripeness detection.

Berry ID	Mass (g)	Shininess	Colour Uniformity	Firmness	Skin strength	Flavor	Sweetness	Texture	Human Ripeness	Target	Machine Ripeness	Confidence (%)
3D9262	8.6849	5	4	4	N	2	3	3	2.5	1	0	2.49
976122	6.4945	4	5	4	N	4	4	3	2.5	1	0	1.09
F3B65 F	7.9683	3	4	4	N	5	5	4	2	1	1	78.48
061AE0	4.9857	5	4	5	N	–	–	–	2	1	1	95.72
7DE558	6.6189	4	5	4	N	4	4	4	3	0	1	96.85
A2761F	7.0942	4	5	4	Y	3	3	3	3	0	0	0.19
8B49E5	8.1658	2	4	2	N	2	3	2	3	0	0	0.01
165B41	6.3459	3	1	4	N	5	5	5	1.5	1	1	99.99
FF9EA0	5.9842	4	1	4	N	5	5	5	1.5	1	1	99.86
E206D2	11.2435	4	5	3	N	2	2	2	3.5	0	0	0.08

**Fig. 10.** Qualitative analysis of the confidence of the machine sensory and human sensory using Pearson correlation matrix.

accuracy on unseen sets and 90.2 % accuracy with increased environmental challenging factors. These results are an 11.3 % and 16.2 % increase in accuracy compared to the RGB-based model. Further experiments compared machine ripeness sensing capability with the human sensory ability, and it was interesting to observe that the machine sensory was positively correlated to the human sensory evaluation of ripeness over blackberry fruit. It was also efficient in overcoming the fuzzy limitations of the human sensory. However, since the hyperspectral imaging in this paper was laboratory-driven, the variability in spectral reflectance due to environmental factors was not considered. Meanwhile, it will be worth investigating the interaction of the environmental factors on reflectance spectra in on-farm settings. Further consideration for future work includes enhancing the stereo camera lens based on the outcome of the on-farm hyperspectral imaging and variability impact analysis. This will likely increase the possibilities of generating numerous samples of better-quality images in a more efficient way, thereby increases the robustness of the proposed model to on-farm spectral variability. There is also room to expand the samples to include the overripe cases which extends the work beyond a binary classification task. With an end goal to develop an ergonomic non-

destructive tool for on-farm fruit ripeness assessment, the study will investigate the applicability of the proposed system on a robot picker, or augmented reality headset for human pickers.

Funding

This work was supported by Innovate UK (50877). The funding sources had no role in the study design, execution, analysis and interpretation of the data, or writing of this manuscript.

CRedit authorship contribution statement

Chollette C. Olisah: Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ben Trehwella:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization. **Bo Li:** Writing – review & editing, Investigation, Formal analysis, Data curation. **Melvyn L. Smith:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Benjamin**

Winstone: Writing – review & editing, Investigation, Formal analysis, Data curation. **E. Charles Whitfield:** Writing – review & editing, Resources, Investigation, Formal analysis, Data curation. **Felicidad Fernández Fernández:** Writing – review & editing, Investigation, Formal analysis, Data curation. **Harriet Duncalfe:** Writing – review & editing, Resources, Data curation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The authors declare that there is no conflict of interest.

Ben Trehwella reports financial support was provided by Innovate UK. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Special thanks to Windmill Hill Fruits Ltd in Herefordshire, UK, and Clockhouse Farm in Kent, UK., for granting access to their farms and permitting the sampling of their blackberry fruits.

References

- Agarap, A.F., 2018. Deep Learning Using Rectified Linear Units (Relu) arXiv preprint arXiv:1803.08375.
- Beghi, R., Giovenzana, V., Spinardi, A., Guidetti, R., Bodria, L., Oberti, R., 2013. Derivation of a blueberry ripeness index with a view to a low-cost, handheld optical sensing device for supporting harvest decisions. *Trans. ASABE (Am. Soc. Agric. Biol. Eng.)* 56 (4), 1551–1559. <https://doi.org/10.13031/trans.56.10169>.
- Castro, W., Oblitas, J., De-La-Torre, M., Cotrina, C., Bazán, K., Avila-George, H., 2019. Classification of cape gooseberry fruit according to its level of ripeness using machine learning techniques and different colour spaces. *IEEE Access* 7, 27389–27400. <https://doi.org/10.1109/ACCESS.2019.2898223>.
- Cho, B.H., Koyama, K., Olivares Díaz, E., Koseki, S., 2020. Determination of “Hass” avocado ripeness during storage based on smartphone image and machine learning model. *Food Bioprocess Technol.* 13 (9), 1579–1587. <https://doi.org/10.1007/s11947-020-02494-x>.
- Chollet, F., 2017. Xception: deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251–1258. <https://doi.org/10.1109/CVPR.2017.195>.
- Cockerton, H., Unzueta, M.B., Johnson, A.V., Diegues, A.V., Fernández, F.F., 2020. Pathway analysis to determine factors contributing to overall quality scores in four berry crops. *J. Hortic. Res.* 21 (1), 35. <https://doi.org/10.2478/johr-2020-0025>.
- Duchi, J., Hazan, E., Singer, Y., 2011. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.* 12 (7).
- Efron, B., 1992. Bootstrap methods: another look at the jackknife. In: *Breakthroughs in Statistics*. Springer, New York, NY, pp. 569–593. <https://doi.org/10.1214/aos/1176344552>.
- El-Bendary, N., El Hariri, E., Hassanien, A.E., Badr, A., 2015. Using machine learning techniques for evaluating tomato ripeness. *Expert Syst. Appl.* 42 (4), 1892–1905. <https://doi.org/10.1016/j.eswa.2014.09.057>.
- Gao, Z., Shao, Y., Xuan, G., Wang, Y., Liu, Y., Han, X., 2020. Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. *Artificial Intelligence in Agriculture* 4, 31–38. <https://doi.org/10.1016/j.aiia.2020.04.003>.
- Gutiérrez, S., Wendel, A., Underwood, J., 2019. Spectral filter design based on in-field hyperspectral imaging and machine learning for mango ripeness estimation. *Comput. Electron. Agric.* 164, 104890 <https://doi.org/10.1016/j.compag.2019.104890>.
- Hamza, R., Chtourou, M., 2018. Apple ripeness estimation using artificial neural network. July. In: *2018 IEEE International Conference on High Performance Computing & Simulation (HPCS)*, pp. 229–234. <https://doi.org/10.1109/HPCS.2018.00049>.
- Hong, D., Yokoya, N., Chanussot, J., Zhu, X.X., 2018. An augmented linear mixing model to address spectral variability for hyperspectral unmixing. *IEEE Trans. Image Process.* 28 (4), 1923–1938.
- Hong, D., Zhang, B., Li, X., Li, Y., Li, C., Yao, J., Yokoya, N., Li, H., Jia, X., Plaza, A., et al., 2023. SpectralGPT: Spectral Foundation Model, p. 2023 arXiv preprint arXiv:2311.07113.
- Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications arXiv preprint arXiv:1704.04861.
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708. <https://doi.org/10.1109/CVPR.2017.243>. Bookmark:
- Kangune, K., Kulkarni, V., Kosamkar, P., 2019. Grapes ripeness estimation using convolutional neural network and support vector machine. In: *2019 IEEE Global Conference for Advancement in Technology (GCAT)*, pp. 1–5. <https://doi.org/10.1109/GCAT47503.2019.8978341>.
- Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization arXiv preprint arXiv:1412.6980.
- Li, B., Cobo-Medina, M., Lecourt, J., Harrison, N., Harrison, R.J., Cross, J.V., 2018. Application of hyperspectral imaging for nondestructive measurement of plum quality attributes. *Postharvest Biol. Technol.* 141, 8–15. <https://doi.org/10.1016/j.postharvbio.2018.03.008>.
- Liu, C., Liu, W., Lu, X., Ma, F., Chen, W., Yang, J., Zheng, L., 2014. Application of multispectral imaging to determine quality attributes and ripeness stage in strawberry fruit. *PLoS One* 9 (2), e87818.
- Liu, W., Zhang, M., Luo, Z., Cai, Y., 2017. An ensemble deep learning method for vehicle type classification on visual traffic surveillance sensors. *IEEE Access* 5, 24417–24425.
- Menardi, G., Torelli, N., 2014. Training and assessing classification rules with imbalanced data. *Data Min. Knowl. Discov.* 28 (1), 92–122.
- Mikulic-Petkovsek, M., Veberic, R., Hudina, M., Zorenc, Z., Koron, D., Senica, M., 2021. Fruit quality characteristics and biochemical composition of fully ripe blackberries harvested at different times. *Foods* 10 (7), 1581.
- Miragaia, R., Chávez, F., Díaz, J., Vivas, A., Prieto, M.H., Moñino, M.J., 2021. Plum ripeness analysis in real environments using deep learning with convolutional neural networks. *Agronomy* 11 (11), 2353.
- Nilsson, N.J., 1965. *Learning Machines Foundations of Trainable Pattern Classifying Systems*. McGraw-Hill, New York.
- Olisah, C.C., Smith, L., 2019. Understanding unconventional preprocessors in deep convolutional neural networks for face identification. *SN Appl. Sci.* 1 (11), 1511. <https://doi.org/10.1007/s42452-019-1538-5>.
- Polder, G., van der Heijden, G.W., Young, I.T., 2002. Spectral image analysis for measuring ripeness of tomatoes. *Trans. ASABE (Am. Soc. Agric. Biol. Eng.)* 45 (4), 1155.
- Raj, R., Cosgun, A., Kulić, D., 2022. Strawberry water content estimation and ripeness classification using hyperspectral sensing. *Agronomy* 12 (2), 425.
- Sabzi, S., Abbaspour-Gilande, Y., García-Mateos, G., Ruiz-Canales, A., Molina-Martínez, J.M., Arribas, J.L., 2019. An automatic non-destructive method for the classification of the ripeness stage of red delicious apples in orchards using aerial video. *Agronomy* 9 (2), 84.
- Saranya, N., Srinivasan, K., Kumar, S.K., 2021. Banana ripeness stage identification: a deep learning approach. *J. Ambient Intell. Hum. Comput.* 1–7.
- Sewell, M., 2008. *Ensemble Learning*. <http://www.machine-learning.martinsewell.com/ensembles/ensemble-learning.pdf>. Accessed July 2022.
- Shah, S.S.A., Zeb, A., Qureshi, W.S., Arslan, M., Malik, A.U., Alasmay, W., Alanazi, E., 2020. Towards fruit maturity estimation using NIR spectroscopy. *Infrared Phys. Technol.* 111, 103479.
- Shahhosseini, M., Hu, G., Khaki, S., Archontoulis, S.V., 2021. Corn yield prediction with ensemble CNN-DNN. *Front. Plant Sci.* 12, 709008.
- Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition arXiv preprint arXiv:1409.1556.
- Sinelli, N., Spinardi, A., Di Egidio, V., Mignani, I., Casiraghi, E., 2008. Evaluation of quality and nutraceutical content of blueberries (*Vaccinium corymbosum* L.) by near and mid-infrared spectroscopy. *Postharvest Biol. Technol.* 50 (1), 31–36.
- Steinbrener, J., Posch, K., Leitner, R., 2019. Hyperspectral fruit and vegetable classification using convolutional neural networks. *Comput. Electron. Agric.* 162, 364–372.
- Sutskever, I., Martens, J., Dahl, G., Hinton, G., 2013. On the importance of initialization and momentum in deep learning. In: *International Conference on Machine Learning*, pp. 1139–1147.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826.
- Wolpert, D.H., 1992. Stacked generalization. *Neural Network* 5 (2), 241–259.
- Worasawate, D., Sakunasinha, P., Chiangga, S., 2022. Automatic classification of the ripeness stage of mango fruit using machine a learning approach. *AgriEngineering* 4 (1), 32–47.
- Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? *Adv. Neural Inf. Process. Syst.* 27.
- Zanini, S., Marzotto, M., Giovannazzo, F., Bassi, C., Bellavite, P., 2015. Effects of dietary components on cancer of the digestive system. *Crit. Rev. Food Sci. Nutr.* 55 (13), 1870–1885.