Nondestructive perception of potato quality in actual online production based on cross-modal technology

Qiquan Wei[1†], Yurui Zheng[1†], Zhaoqing Chen[2], Yun Huang[3], Changqing Chen[3], Zhenbo Wei[4], Shuiqin Zhou[5], Hongwei Sun[1], Fengnong Chen[1*]

(1. *School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China*;
2. *School of Information Engineering and Internet of Things, Huzhou Vocational and Technical College, Huzhou 313000, Zhejiang, China*;
3. *Jinhua Academy of Agricultural Sciences, Jinhua 321017, Zhejiang, China*;
4. *College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China*;
5. *Fair Friend Institute of Intelligent Manufacturing, Hangzhou Vocational and Technical College, Hangzhou 310018, China*)

Abstract: Nowadays, China stands as the global leader in terms of potato planting area and total potato production. The rapid and nondestructive detection of the potato quality before processing is of great significance in promoting rural revitalization and augmenting farmers' income. However, existing potato quality sorting methods are primarily confined to theoretical research, and the market lacks an integrated intelligent detection system. Therefore, there is an urgent need for a post-harvest potato detection method adapted to the actual production needs. The study proposes a potato quality sorting method based on cross-modal technology. First, an industrial camera obtains image information for external quality detection. A model using the YOLOv5s algorithm to detect external green-skinned, germinated, rot and mechanical damage defects. VIS/NIR spectroscopy is used to obtain spectral information for internal quality detection. A convolutional neural network (CNN) algorithm is used to detect internal blackheart disease defects. The mean average precision (mAP) of the external detection model is 0.892 when intersection of union (IoU) = 0.5. The accuracy of the internal detection model is 98.2%. The real-time dynamic defect detection rate for the final online detection system is 91.3%, and the average detection time is 350 ms per potato. In contrast to samples collected in an ideal laboratory setting for analysis, the dynamic detection results of this study are more applicable based on a real-time online working environment. It also provides a valuable reference for the subsequent online quality testing of similar agricultural products.

Keywords: cross-modal technology, potato quality, YOLOv5s, VIS/NIR spectroscopy, online nondestructive detection
DOI: 10.25165/j.ijabe.20231606.8076

Citation: Wei Q Q, Zheng Y R, Chen Z Q, Huang Y, Chen C Q, Wei Z B, et al. Nondestructive perception of potato quality in actual online production based on cross-modal technology. Int J Agric & Biol Eng, 2023; 16(6): 280–290.

1 Introduction

Potatoes rank as the fourth most significant global food crop in terms of production volume[1]. In 2020, China's potato production is 130 million t, accounting for 24.91% of global production. Potatoes require 30% less water for cultivation compared to rice, yet exhibiting higher yields per hectare. Additionally, potatoes are well-suited to thrive in adverse environments, including arid and alpine regions. However, the presence of defects such as mechanical damage, rot, germination, green-skin and internal black heart disease can significantly diminish the economic value of potatoes. Therefore, the development of intelligent testing equipment capable of comprehensively detecting both internal and external defects in potatoes holds immense significance in safeguarding food security, alleviating resources and environment pressures and promoting farmers' income and agricultural development in China.

Nondestructive detection technology has gained widespread utilization in agricultural products detection, owing to its advantages of high repeatability, high detection efficiency compared with the traditional manual detection techniques. Machine vision technology is widely used in the external nondestructive detection of agricultural products. For instance, Jin et al.[2] used area thresholding and black ratio thresholding methods to identify defects in potatoes, with better results for mechanical damaged defects. Ebrahimi et al.[3] identified the green-skinned defects of potatoes through color cameras and compared the characteristics of different bands of RGB for defects identification. Barnes et al.[4] employed machine learning algorithm to identify potato defects with an accuracy rate of over 89.5%. Elmasry et al.[5] classified irregular potatoes to sort out malformed potatoes. With the increasing advancements in deep learning research, scholars have started using deep learning techniques for defect detection in agricultural products. Oppenheim et al.[6] developed a CNN model to classify four categories of potato disease defects. Elsharif et al.[7]

Received date: 2022-12-11 **Accepted date:** 2023-05-21
Biographies: Qiquan Wei, MS candidate, research interest: intelligent agricultural equipment, Email: 222060278@hdu.edu.cn; **Yurui Zheng**, MS candidate, research interest: intelligent agricultural equipment, Email: 212060336@hdu.edu.cn; **Zhaoqing Chen**, MS, research interest: intelligent agricultural equipment, Email: 454507269@qq.com; **Yun Huang**, Agronomist, research interest: intelligent agricultural equipment, Email: 26185188@qq.com; **Changqing Chen**, Agronomist, research interest: intelligent agricultural equipment, Email: 496533235@qq.com; **Zhenbo Wei**, PhD, Professor, research interest: development and application of intelligent sensory system, Email: weizhb@zju.edu.cn; **Shuiqin Zhou**, PhD, Associate Professor, research interest: nondestructive testing of agricultural products, Email: 2004010033@hzvtc.edu.cn; **Hongwei Sun**, PhD, research interest: intelligent agricultural equipment, Email: sunhongwei@hdu.edu.cn.
†These authors contributed equally to this work.
***Corresponding author: Fengnong Chen**, PhD, Associate Professor, research interest: intelligent agricultural equipment. Hangzhou Dianzi University, Hangzhou 310018, China. Tel: +86-13588805823. Email: fnchen2022@126.com.

focused on classifying potatoes of different colors (red, red washed, sweet, and white). Chen et al.[8] used migration learning for rice disease detection, achieving an average accuracy of 98.63% for identification. Zhao et al.[9] used CNN models to classify six types of soybean seeds and the classification accuracy was 98.87%. Ramos et al.[10] used the VGG19 network model to classify the grape ripeness，achieving a maximum accuracy of 93.41%. Additionally, machines for sorting garlic[11], pistachios[12] and pomegranate seeds[13] are currently available. Most scholars use classification models to classify agricultural products' shape defects and colors. However, in defect detection, the classification model can only identify the class of the whole image, and may not accurately identify the specific location of small defects, leading to unrecognized situations. On the other hand, employing object detection methods, such as using anchor boxes, can accurately locate the location of defects, which can improve the accuracy of defect detection in potatoes.

Many researchers have also conducted in-depth research on internal nondestructive detection of agricultural products. Hajjar et al.[14] used MRI to study internal defects in potato tubers. Sosa et al.[15] used X-ray fluorescence spectrometry to estimate iron and zinc concentrations in potato tubers. Hyperspectral imaging technology has also been used by several scholars for defect detection in agricultural products[16-18]. However, MRI and X-ray technology are associated with high costs, and hyperspectral imaging technology requires a large amount of data and a long calculation time, which is difficult to use for online detection. In contrast, VIS/NIR spectroscopy has emerged as a promising technique for nondestructive detection of agricultural products, such as oranges[19], apples[20], root vegetables[21], lemons[22], pears[23], and drupe[24]. VIS/NIR spectroscopy offers advantages such as fast detection speed and minimal equipment requirements, making it well-suited for potato internal quality detection[25-27]. For instance, Zhou et al.[28] investigated the feasibility of using VIS/NIR transmission spectroscopy in the 513-850 nm region combined with PLS-LDA for the classification of blackhearted potatoes, achieving an accuracy of 96.53% in the test set. However, it should be noted that the accuracy was determined in a laboratory setting and serves as a reference for online detection.

The above analysis indicates that machine vision technology is effective in detecting the external quality of potatoes but cannot reflect the internal quality information. VIS/NIR spectroscopy technology can accurately characterize the internal quality of potatoes but may have limitations in reflecting the external quality. To overcome these limitations, researchers have studied Data fusion techniques by combining image-based and spectroscopy-based approaches. Song et al.[29] utilized near-infrared spectroscopy and computer vision systems to evaluate the flavor and appearance quality of tea. The accuracy of the test set increased from 98.57% to 100% when features from multiple sensors were used. Yu et al.[30] used machine vision technology to locate wood defects and used near-infrared spectroscopy to obtain spectra data for defect detection. Furthermore, image-spectroscopy fusion has been explored in other applications such as fish freshness detection[31] and foreign contaminants detection in food[32].

Therefore, it is feasible to design a method for potato defect detection by integrating machine vision and VIS/NIR spectroscopy. This study employs deep learning and spectral analysis to achieve rapid and nondestructive detection of the external quality problems such as green-skin, germination, rot, mechanical damage, as well as internal quality problems like blackheart disease in potatoes. In this work, an online flow line sorting system is developed to efficiently eliminate potato defects.

The organization of this paper is outlined as follows: The first part discusses the methods used to achieve nondestructive detection of internal and external potato defects. The second part introduces the design of hardware system, the acquisition of data set, the utilization of relevant algorithms for internal and external potato defect detection and the establishment of verification standards. The third part presents the results of the algorithm experiments and their application in the actual production line. In the algorithm experiment part, various comparative experiments are conducted, and the results are comprehensively evaluated and analyzed. Lastly, a summary is presented in the fourth part.

2 Materials and methods

2.1 Experimental samples

2.1.1 Defective samples

Potatoes are susceptible to various forms of damage during their growth and harvesting stages, leading to significant degradation in fruit quality. The experimental potato samples utilized in this study were purchased from local farmers' markets. The selection of five specific types of defects for this experiment is in accordance with relevant national industry standards[33]. Figure 1 shows the sample of these defects. It is worth noting that the sample of different degrees of black heart disease are cut for display and whole potatoes are used in actual online production. Defects are defined as follows.

Germinated defects: Germinated defects are characterized by the presence of visible germs on the surface of potatoes, which can result in the production of solanine, a harmful compound for humans.

Mechanical damage defects: Mechanical damage defects occur due to deep cracks in the flesh of potatoes caused by external mechanical stimulation during harvesting, increasing the risk of internal tissue infection.

Rot defects: Rot defects manifest as areas of wet rot on the potato surface following mechanical damage.

Green-skinned defects: Green-skinned defects are caused by exposure of potatoes to sunlight, diffused light, or other sources of light, resulting in the development of green coloration in the epidermis and flesh.

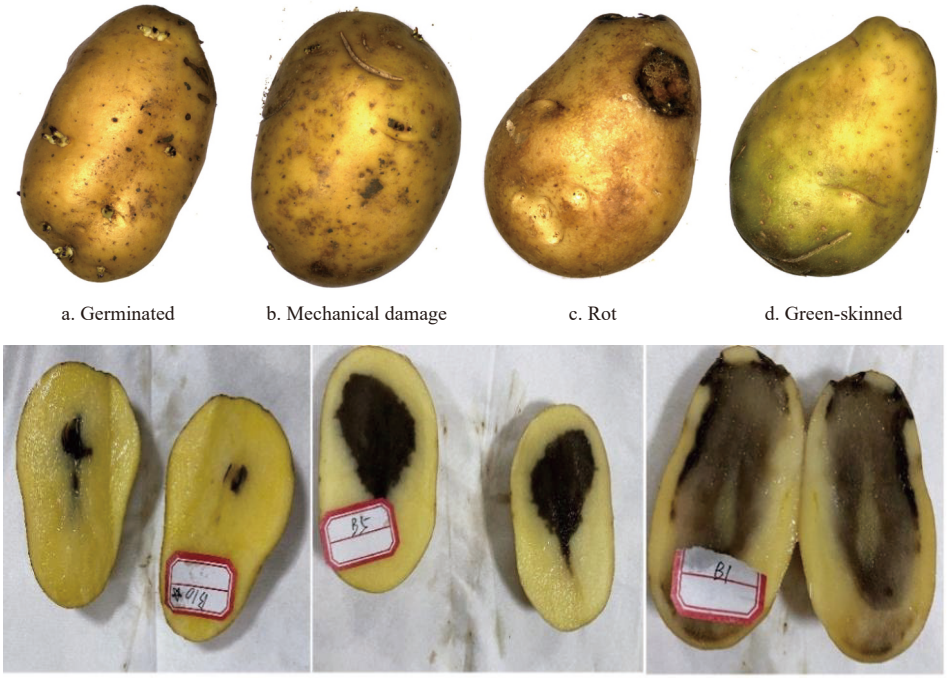
Internal blackheart disease defects: Internal blackheart disease defects occur due to anaerobic respiration in an oxygen-deficient environment, leading to internal tissue rot and blackening.

To augment the availability of samples for the study, artificial methods were employed to induce certain defects in potatoes that may be difficult to procure directly from the market. For example, green-skinned defects were obtained by exposing potatoes to light. After the surface of these potatoes were intentionally damaged, they were placed in the black plastic bags and cultured for several days in a humid environment at room temperature to induce germination and rot. To simulate internal defects, potatoes were washed and individually sealed in plastic bags and then subjected to a temperature of 38°C-40°C. After 24-36 h, symptoms of slight blackheart disease appeared while moderate and severe blackheart disease potatoes were cultivated for 48-72 h. This approach facilitated the acquisition of samples with varying degrees of blackheart disease, thereby expanding the dataset for the study.

2.1.2 Samples collection platform

As is shown in Figures 2a and 2b, a comprehensive data acquisition setup comprising of an image data acquisition platform and a spectral data acquisition platform was established to capture

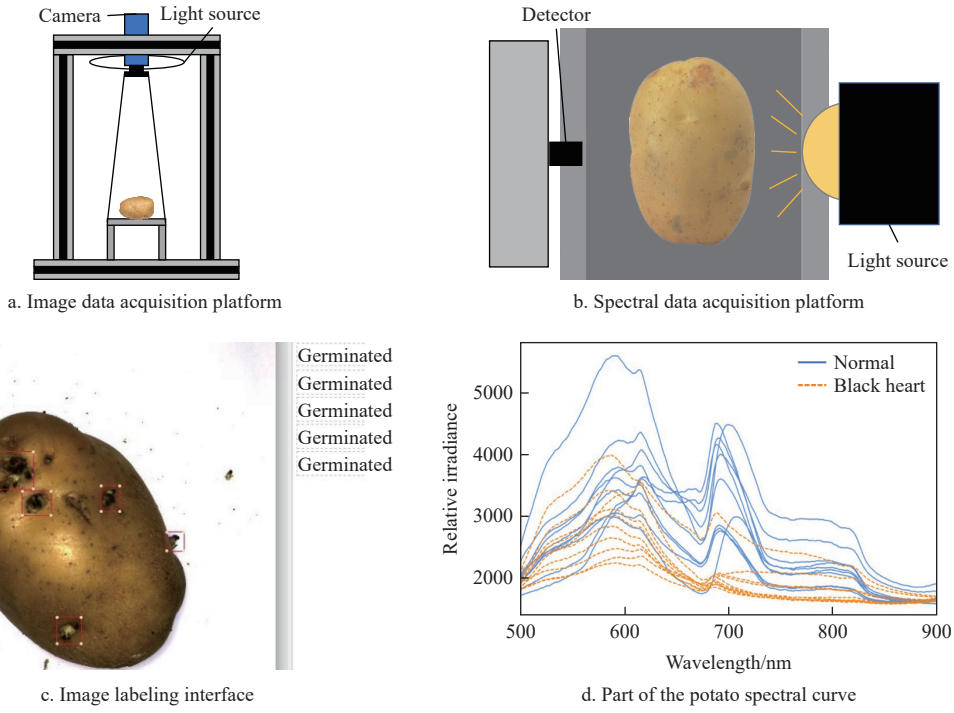
external image data and internal spectral data of potatoes. The image data acquisition system uses an industrial camera (MV-CE050-30UC, HIKVISION, Hangzhou, China) with a 16 mm fixed-focus lens and white LED ring light source (DRL-170-45-W, HZN Automation Technology Co., Shanghai, China). The spectral data acquisition system uses a spectrometer with a slit width of 100 μm (FLA5000, Fight Technology Co., Hangzhou, China) and a 600 μm core diameter of fiber optic (FIB-600-NIR, IdeaOptics Technologies, Shanghai, China) to measure the spectral range of 200-1100 nm. A halogen lamp is used as a light source (50 W, Philips Lighting NV, Shanghai, China). In Figure 2a, a camera is positioned on top of the potato to acquire image data of the upper surface. In Figure 2b, the fiber optic is placed on opposite side to the spectrum to obtain the VIS/NIR transmission spectrum of the potato.



a. Germinated b. Mechanical damage c. Rot d. Green-skinned

e. Different degrees of blackheart disease

Figure 1 Schematic diagram of defective potato samples



a. Image data acquisition platform b. Spectral data acquisition platform

c. Image labeling interface d. Part of the potato spectral curve

Figure 2 Data acquisition platform and building datasets

2.1.3 Potatoes data collection

To enhance the robustness of the model, the following three considerations are considered during the production of the potato image dataset:

1) Color features. The surfaces of defective potatoes exhibit distinct colors, with green-skinned defects appearing as green, blue-

green, or gray-green, rot defects appearing as gray or gray-black and germinated defects appearing as white or dark-green. Therefore, accurate color representation is crucial for locating and identifying potato defects. To ensure color fidelity, the white balance of the industrial camera was calibrated prior to image acquisition.

2) Distance impact. The shooting distance can significantly affect the size of potatoes in the acquired images, as well as the size of the defects in the images. This is particularly relevant for small defects such as germinated defects. To account for this, the height of the camera holder is adjusted to obtain potato images from different shooting distances.

3) Light impact. Different illumination conditions can result in variations in the imaging effects. Therefore, the dataset is expanded by modifying the exposure time of the industrial camera to simulate different lighting conditions, thereby adapting the model to more lighting conditions.

The experiment involved 2307 images of potatoes, each of which was meticulously annotated with rectangular bounding boxes to delineate the areas of defects using the Colabeler labeling tool as exemplified in Figure 2c. The defects were comprehensively labeled, with a total of 2389 defects labeled as germinated defects, accounting for 53.2% of the dataset; 700 defects are labeled as mechanical damage, accounting for 15.6% of the dataset; 776 defects are labeled as rot, accounting for 17.3% of the dataset; 623 instances are labeled as green-skinned, accounting for 13.9% of the dataset.

The spectrometer was powered on and allowed to preheat for a duration of 30 min to ensure stable performance. Subsequently, the original spectral data was acquired using the spectrometer, with the integral time set to 80 ms and the number of spectral smoothing frequency set to 3. To reduce interference in the data, the acquisition wave band is carefully chosen as 500-900 nm[34]. Spectral data is collected at different locations and angles of potatoes, and 280 samples of normal potatoes and 280 samples of black heart potatoes are collected. Representative spectral data points are illustrated in Figure 2d, with wavelength depicted on the horizontal axis and relative irradiance on the vertical axis. The spectral data of potatoes containing blackheart disease is depicted by the yellow curve, while the spectral data of normal potatoes is represented by the blue curve. Notably, discernible differences between the spectral data of potatoes with internal defects and healthy potatoes can be observed around 700 nm. The acquired spectral data, along with their corresponding class labels, were compiled to create a comprehensive dataset for subsequent development of a non-destructive detection model.

2.2 Relative algorithms

2.2.1 External detection algorithm

The YOLOv5 object detection algorithm is employed to detect external defects in potato images. The fundamental concept of the YOLO framework involves taking the entire image grid as input and use a convolutional network to obtain bounding box confidence and class probability of multiple candidate bounding box. Currently, YOLO series models have progressed from YOLOv1 (2016) to YOLOv5 (2020). YOLOv5 boasts several advantages, including high detection accuracy, efficient detection speed and a compact model size with fewer parameters. Compared to YOLOv3 and YOLOv4, YOLOv5 has been demonstrated to achieve superior accuracy in detecting mold and other defects on food surfaces[35].

The YOLOv5 network architecture comprises input, backbone, neck network and head components, as illustrated in Figure 3. On the input terminal, YOLOv5 employs mosaic data augmentation, adaptive anchoring and adaptive image scaling techniques. Mosaic data augmentation enhances the detection performance of small objects. The adaptive anchor frame method adaptively calculates the optimal anchor frame values for different training sets during training and adaptive image scaling reduces the computational effort of model inference, resulting in a significant improvement in detection speed. The backbone network layer is a convolutional neural network that aggregates and extracts image features at different fine-grained levels, composed of FOCUS, CBL, CSP1 X and other modules. The neck network layer mainly mixes and combines a series of image features and passes them to the prediction layer, which consists modules of CBL, upsampling, CSP2 X and other modules. The head terminal mainly generates the

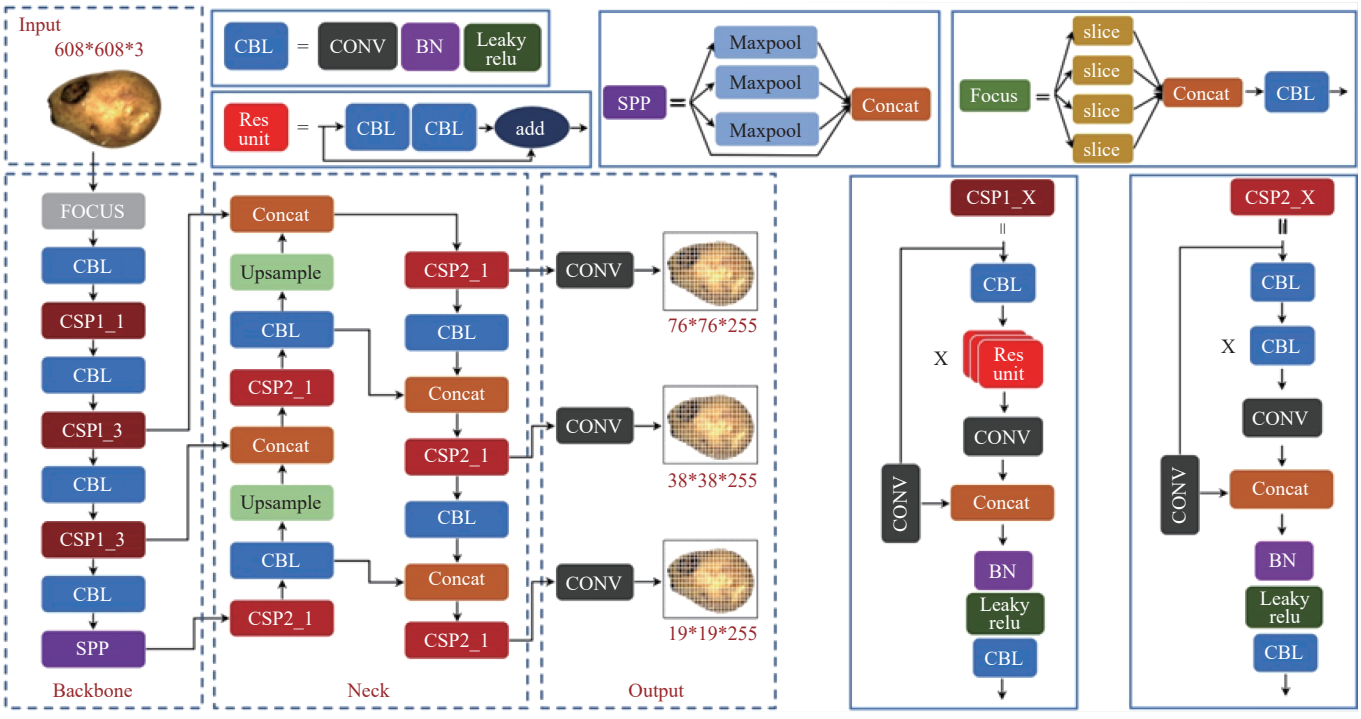


Figure 3 YOLOv5 network structure diagram

bounding box and predicts the category for the image features, and this part uses GIoU Loss as the loss function of the bounding box.

YOLOv5 include several variants, including YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. These variants share a consistent backbone, neck, and head architecture, with the main differences lying in the number of feature extraction modules and convolutional sums at specific locations (Table 1). Due to the increase in the number of feature extraction modules and convolution sums, YOLOv5l and YOLOv5x have more parameters. Although the detection accuracy of them is improved, the detection speed will be affected in the online detection. YOLOv5s and YOLOv5m require fewer model parameters, resulting in faster algorithm execution during actual detection, making them more suitable for real-world scenarios. The model accuracy can be further improved with an increase in the subsequent sample dataset, compensating for any potential loss in choosing a smaller parameter model. Considering this, this study employs YOLOv5s and YOLOv5m models for comparison of results.

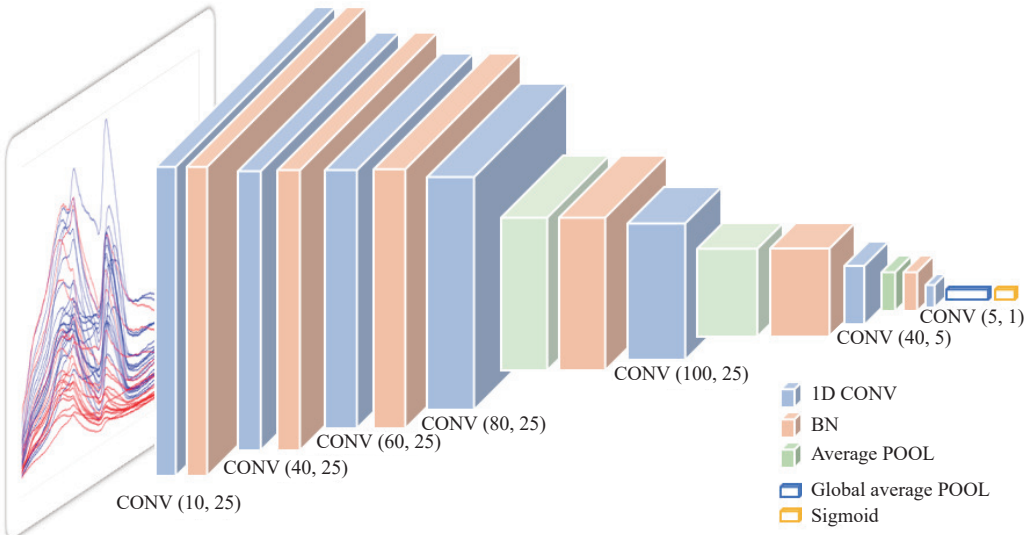
Table 1 Comparison of different YOLOv5 model network depths

Model	Backbone: CPS1_X			Neck: CSP2_X				
	First	Second	Third	First	Second	Third	Fourth	Fifth
YOLOv5s	CSP1_1	CSP1_3	CSP1_3	CSP2_1	CSP2_1	CSP2_1	CSP2_1	CSP2_1
YOLOv5m	CSP1_2	CSP1_6	CSP1_6	CSP2_2	CSP2_2	CSP2_2	CSP2_2	CSP2_2
YOLOv5l	CSP1_3	CSP1_9	CSP1_9	CSP2_3	CSP2_3	CSP2_3	CSP2_3	CSP2_3
YOLOv5x	CSP1_4	CSP1_12	CSP1_12	CSP2_4	CSP2_4	CSP2_4	CSP2_4	CSP2_4

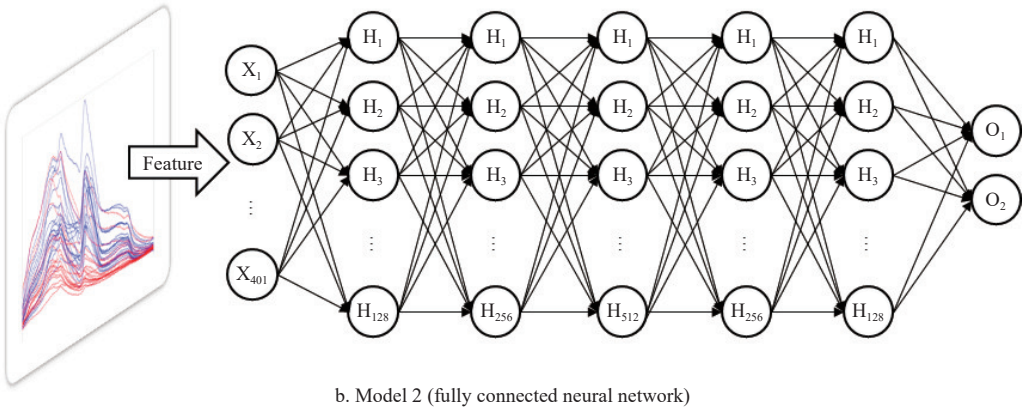
2.2.2 Internal detection algorithm

In the context of detecting internal potato defects, deep learning algorithms are employed due to their ability to automatically learn data features from large inputs and their computational efficiency.

In this study, the detection of internal defects is formulated into a binary classification problem and two models, Model 1 and Model 2, are developed for comparison: Model 1 is a one-dimensional convolutional neural network(1D-Conv), inspired by the work of Rong et al.[36] and Model 2 is a fully connected neural network (FCN). The structure diagram of these models is shown in Figure 4. The 1D-Conv model incorporates three convolutional layers with batch normalization (BN) layers to mitigate gradient vanishing and three blocks (1D-Conv average pool, BN) to extract deep feature from the input spectral variables, which are one-dimensional features. The classifier output unit, consisting of global average pooling and sigmoid layers, is used for the identification of internal defects. Model 2 consists of an input layer, multiple hidden layers and an output layer with sigmoid activation functions. Both of two models use the Adamax optimizer.



a. Model 1 (convolutional neural network)



b. Model 2 (fully connected neural network)

Figure 4 Internal model structure diagram

2.3 Hardware and software design

2.3.1 Hardware system

To be close to a real-world production environment, an industrialized flow line is built. The system consists of a transport conveyor, a spectral data acquisition system, an image data acquisition system and a sorting system. The hardware setup used in the study is consistent with the description in Section 2.1.2. The operational process of the hardware system is shown in Figure 5a. Potatoes are placed on pallets and conveyed along the flow line when the system is running. The camera and spectrometer will be triggered by infrared sensors when the potato reaches the designated location. Then the software system obtains the image and spectral information and employs the developed algorithms to analyze the potatoes for the presence of defects. Based on the determination, the potatoes are sorted into different channels.

2.3.2 Software system

The software system is developed using the C++ language and Qt Creator, based on the Qt framework and serves as the control interface for the entire system. OpenCV, an open-source computer vision library (version 4.5.2), is used to facilitate the operations of images[37]. In this study, functions of OpenCV are used for image processing on the acquired images and the deep learning models are loaded using the DNN module. The algorithmic workflow is depicted in Figure 5b. Firstly, the region of interest (ROI) of a potato in the image is identified. If no ROI is found, it indicates that the pallet does not contain a potato, and the model algorithm does not proceed. If an ROI is detected, the trained YOLOv5s model is loaded for image information detection, and the CNN model is loaded for spectral> information detection. Based on the model results, a comprehensive judgment is made to determine the presence of defects.

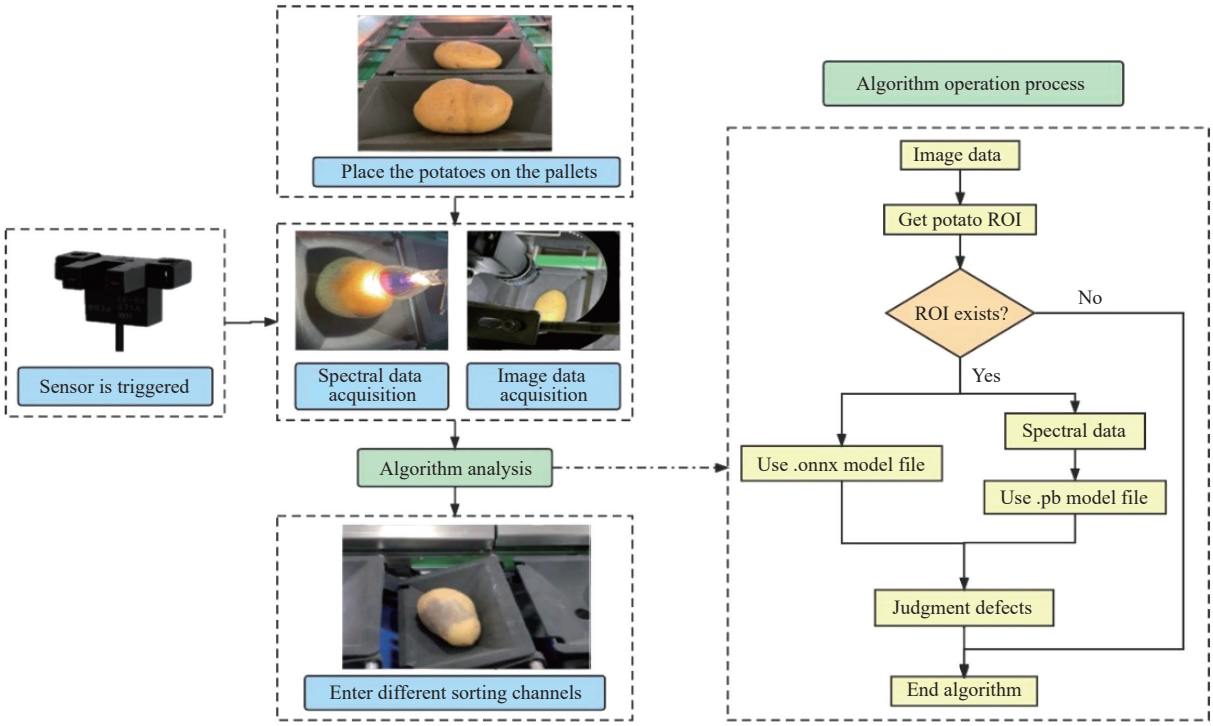


Figure 5 System operation and algorithm process flow

2.4 Experimental setup

2.4.1 Implementation details

A computer with an Intel® CPU i9-10900K @3.7GHz processor, 32GB of RAM, and NVIDIA GeForce RTX 3060 12GB is used in the deep learning network model training of this study. Python3.8 was used as the main programming language. A computer with an AMD Ryzen 9 4900HS @3.00GHz processor, 16GB of RAM, and NVIDIA GeForce RTX 2060 6GB is used to run the software system.

2.4.2 Model evaluation metrics

In order to provide a comprehensive evaluation of the model performance, a confusion matrix is used in this study. TP, FN, FP, and TN indicate true positives, false negatives, false positives, and true negatives, respectively, as listed in Table 2.

Table 2 Confusion matrix

		Predicted value	
		Positive	Negative
True value	Positive	TP	FN
	Negative	FP	TN

In the evaluation of the external defect detection model, the primary evaluation metrics employed are the average precision (AP) and mean average precision (mAP). AP, defined as the area under the curve of the precision-recall curve, is used to measure the overall performance of the model in terms of precision and recall. The mAP is the mean value of each category AP after summation. Intersection over union (IoU) calculates the ratio of the overlap between the predicted box to the true box. A standard IoU threshold of 0.5 is typically used to calculate AP and mAP.

In the evaluation of internal defect detection model, the three key evaluation metrics, accuracy, recall, and F1-score are used. Accuracy measures the overall correctness of the classification; recall is the ratio obtained by dividing the number of correct cases in the sample predicted to be positive by the sample of all positive cases, which measures the classifier's ability to identify positive cases; precision is the proportion of the samples classified as positive cases that are actually positive cases; the F1-score, a widely used metric for classification problems, is the harmonic mean of precision and recall, and higher F1-score indicates better model performance.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{AP} = \int_0^1 P(R)\,\text{d}R \quad (3)$$

$$\text{IoU} = \frac{\text{Detect result} \cap \text{Ground truth}}{\text{Detect result} \cup \text{Ground truth}} \quad (4)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5)$$

$$\text{F1} - \text{sorce} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

3 Results and discussion

3.1 External detection

3.1.1 Overview of image

The acquired image is a three-channel RGB image. As is mentioned in Section 2.1.3, the surfaces of defective potatoes exhibit distinct colors. So the color feature can be used as a basis for detecting defects[38].

3.1.2 Comparison of external detection models

To examine the impact of training data size on model accuracy, data augmentation techniques are used in this study. As shown in Figure 6, a single image is augmented to 10 images with variations. Specifically, Gaussian blur, brightness transformation, mirror flip, left-right flip, zoom in/out, and angle rotation are applied randomly to the images during data augmentation. The performance of the models is then evaluated on the same test set to assess the effectiveness of the data augmentation.

Before training model, the Mosaic data augmentation technique is applied to the dataset in this study. Mosaic data augmentation involves performing translation, rotation, scaling, and other transformations on four images, followed by synthesizing these images into one to increasing the diversity of the data. The blending of multiple images is equivalent to increasing the batch size, which makes the training converge faster. The experimental training parameter epoch is set to 100. The AP@0.5 and mAP@0.5 of the model in the training set are listed in Table 3. The result shows that the models trained with the data augmentation exhibit a 3%-5% improvement in mAP@0.5, compared to the non-augmented models.

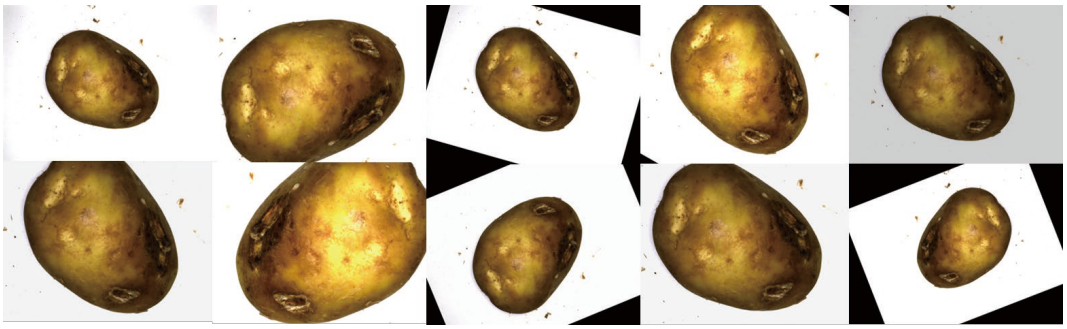


Figure 6 Data augmentation effects

Table 3 Comparison of the performance of the models, YOLOv5s and YOLOv5m, for potatoes defects detection

Data type	Model	Rot AP/%	Green-skinned AP/%	Mechanical damage AP/%	Germinated AP/%	mAP/%	Model size/MB
Unaugmented data	YOLOv5s	86.2	**97.5**	71.0	86.6	84.4	**13.7**
	YOLOv5m	85.7	96.2	75.9	**88.1**	86.5	45.5
Augmented data	YOLOv5s	94.6	90.4	**84.0**	87.9	89.2	**13.7**
	YOLOv5m	**95.4**	93.6	83.2	86.9	**89.8**	45.5

Comparing the model performance revealed that there was no significant difference in terms of AP and mAP between YOLOv5s and YOLOv5m, with values of 89.2% and 89.8%, respectively. However, the model size of YOLOv5m is nearly three times larger than that of YOLOv5s, which suffers greatly in regard to the efficiency of real-time detection. Considering the efficiency influence to the flow line, the YOLOv5s algorithm is ultimately chosen as the model for external defect detection because its small parameters are well suited for application. The current mAP of the model may be limited by the insufficient sample size, and further expansion of the dataset with additional images in subsequent production can effectively improve the mAP of the model.

Figure 7 presents the model's detection results on the potato dataset, showing the detection results of different defect types and sizes. Figure 7a is the result of the unaugmented training dataset and Figure 7b is the result of the augmented training dataset. As the results show, the model trained by augmented dataset can accurately detect potato defects, has higher accuracy for defect detection, and detects defects that are not detected by the original model. The data augmentation method improves the robustness of the model.

Figure 8 shows the AP confusion matrix of the YOLOv5s models' results on the test set. Figure 8a displays the result of the unaugmented training dataset while Figure 8b displays the result of the augmented training dataset. The horizontal axis of the figure shows the predicted results, and the vertical axis shows the actual results, with R indicating rot, GR indicating green-skinned, M indicating mechanical damage, GE indicating germinated, and BG representing background. It is observed that the model trained with the augmented training dataset has significantly improved mAP, albeit with a decrease in AP of green-skinned defects. This could be attributed to the use of brightness transformation during data augmentation, which may impact the detectability of features related to green-skinned defects, consequently leading to a decrease in AP for this defect type.

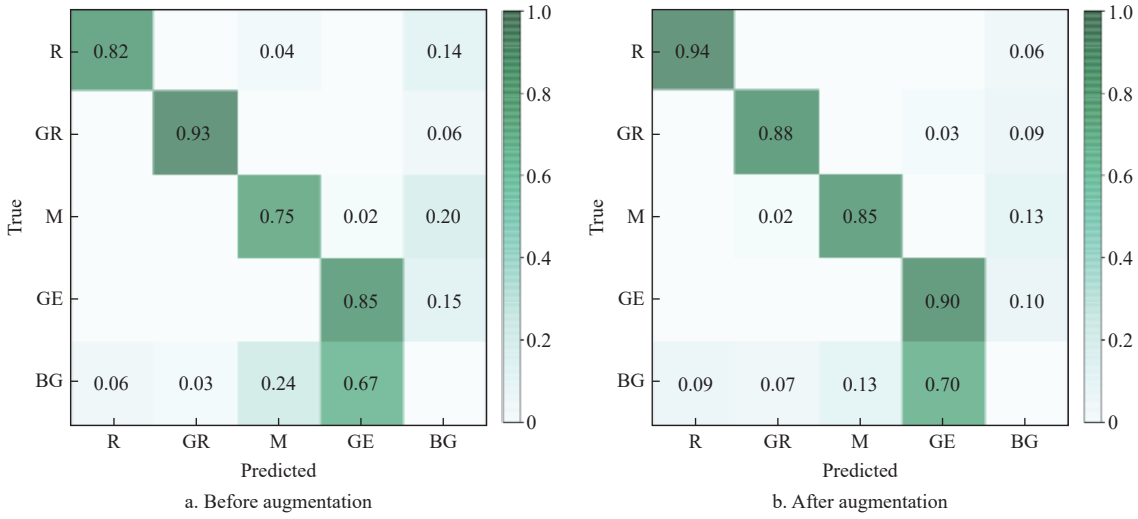
The models misjudge the background seriously. The probability of background misjudgment as germinated was 0.67 and 0.70 in the two models, and the probability of misjudgment as mechanically damaged was 0.24 and 0.13, misjudgment of the background can affect detection accuracy. To alleviate this issue, a two-step approach is employed, wherein the presence of potatoes on a pallet is first determined, followed by algorithmic analysis and processing for flow line detection. Therefore, the maximum contour algorithm

is used to identify the potato region. The source image is converted to HSV color space, and segmentation of the color is performed to extract the potato portions. Subsequently, grayscale and binary processing are applied. Finally, morphological operations are conducted to eliminate small black dots and determine outer contours, the largest contour is identified as the potato ROI. With this algorithm, it is possible to suppress environmental influences and improve detection accuracy.

a. Detectionresult of unaugmented training dataset

b. Detection result of augmented training dataset

Figure 7 Comparison of detection algorithm before and after training dataset augmentation



a. Before augmentation b. After augmentation

Note: R is rot, GR is green-skinned, M is mechanical damage, GE is germinated, and BG is background.

Figure 8 YOLOv5s confusion matrix for the test dataset detection effect before and after training dataset augmentation

3.2 Internal detection

3.2.1 Overview of Spectra

The prominent spectral response observed in the potato samples is characterized by the doubled and combined frequency absorptions of hydrogen-containing groups C-H, O-H, N-H, S-H, and P-H vibrations. Based on the spectral data acquired, it appears that internal blackheart disease tissues of potatoes affect the spectral content transmitted through the potatoes. The energy value of the transmission lowers with the severity of the blackheart disease. The difference in spectral data between the potatoes with slight blackheart disease and normal potatoes is not very significant, this will increase the difficulty of detection. These changes in spectral response may be attributed to the presence of black portions in the flesh, drier (i.e., lower OH) regions, and flour-textured cortical tissue in the diseased potatoes. The drier tissue might typically be expected to increase the apparent absorbance, as light is scattered more due to increased numbers of air tissue interfaces, resulting in reduced transmitted intensity[39].

3.2.2 Comparison of internal detection models

The dataset is divided into training and test sets at a ratio of 8:2. Three metrics, accuracy, recall and F1-score, are used to evaluate the performance of the internal detection models. To assess the performance of algorithm models based on deep learning, this study compares the proposed method with the decision tree (DT), random forest (RF), logistic regression (LR) and support vector machine (SVM) methods. The performance parameters of the models on the test set are listed in Table 4. Model 1 has the highest accuracy at 98.6%, with Model 2 achieving the second-highest accuracy at 95.5%. The DT method shows a poor accuracy rate of only 81.3%. This is caused by the instability of the DT itself, where small changes in the data can lead to completely different tree generations and affect the model performance. In terms of recall and F1-score, Model 1 demonstrated the best performance, achieving scores of 98.2% and 98.2%, respectively, while Model 2 attained scores of

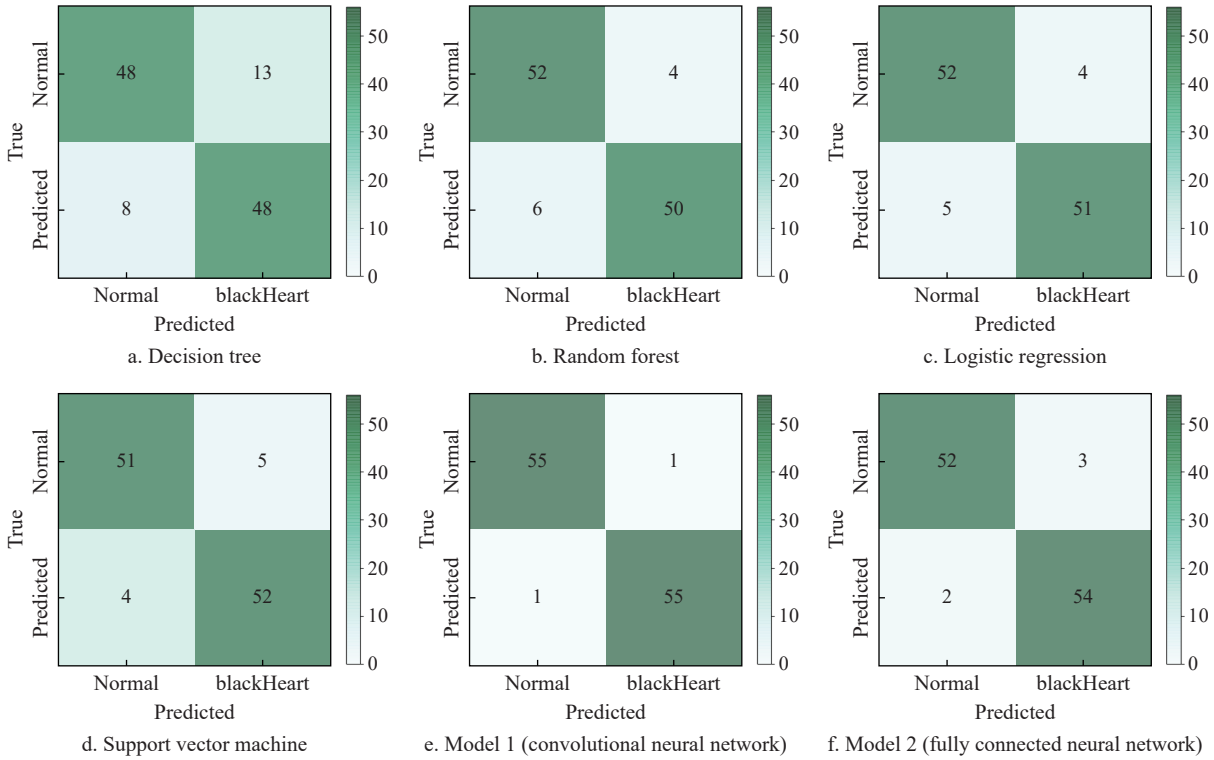
94.6% and 95.4%, respectively, ranking second to Model 1. Based on the above comparison, it can be concluded that the algorithm model builds by the convolutional neural network outperforms the traditional machine learning methods in terms of detection accuracy and overall performance.

Table 4 Model comparison effect on the test set

Model	Accuracy/%	Recall/%	F1-score/%	Model size/MB
DT	81.3	85.7	82.0	0.007
RF	91.0	89.3	90.9	0.46
LR	92.0	91.1	91.9	**0.004**
SVM	92.0	92.9	92.0	0.26
Model 1	**98.2**	**98.2**	**98.2**	1.62
Model 2	95.5	94.6	95.4	1.43

Note: DT is Decision tree, RF is random forest, LR is logistic regression, SVM is support vector machine, Model1 is convolutional neural network and Model2 is fully connected neural network.

The spectral characteristics of potatoes with blackheart disease and normal potatoes exhibit both similarities and differences, manifested in the position and intensity of the spectral feature peaks. For potato spectral data with slight blackheart disease, the spectral feature peaks tend to resemble those of normal potatoes, potentially leading to misclassification by the detection model. As shown in Figure 9, a confusion matrix is constructed to analyze and compare the generalization performance of each model on the test set. The horizontal axis of the figure shows the prediction results, and the vertical axis shows the actual results. Among the 112 test samples, the deep learning-based model demonstrated superior detection performance compared to other methods, with only a very small number of potatoes not being correctly sorted. Model 1 had the best results, with only 1 missed and 1 false detection.



a. Decision tree b. Random forest c. Logistic regression

d. Support vector machine e. Model 1 (convolutional neural network) f. Model 2 (fully connected neural network)

Figure 9 Confusion matrix of the model's detection effect in the test set

In summary, Model 1, the proposed algorithm model built by the convolutional neural network with the best effect, is selected as the model for online internal defect detection in this study. The limited sample size remains a challenge leading to false detections and missed detections, indicating potential for further optimization of the model as the sample size increases.

3.3 System realization

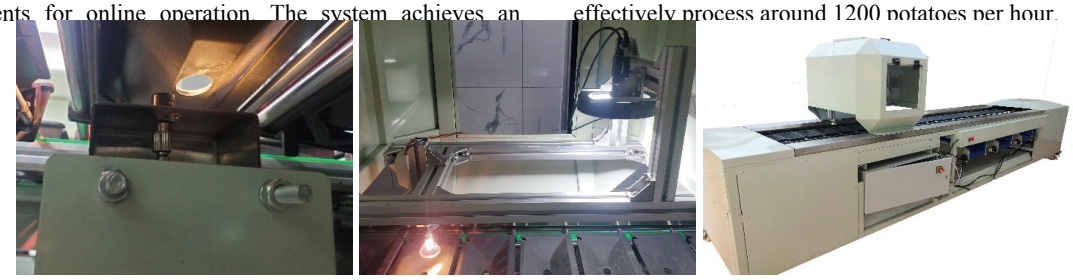
3.3.1 Hardware System

The hardware system is shown in Figure 10. It comprises a halogen lamp positioned 87mm away from the fiber optic, a fiber optic situated 12mm above the pallet, and a vision system positioned 360mm above the pallet. The detection system is placed in a dark box to minimize the impact of external environmental factors. The software system communicates with the hardware system by the Modbus/TCP protocol. Upon clicking the start button, the hardware system receives the start command and begins the operation of the flow line. Each pallet has a small hole at its center, and as the potatoes pass through this hole in the pipeline, the sensor in the hardware system is triggered, sending a command to the software system. After receiving the command, the software system captures the image data from the camera and spectrum data from the spectrometer. The algorithm then analyzes the spectral and image data to detect the presence of defects in the potatoes and classifies them into different channels.

3.3.2 Software system

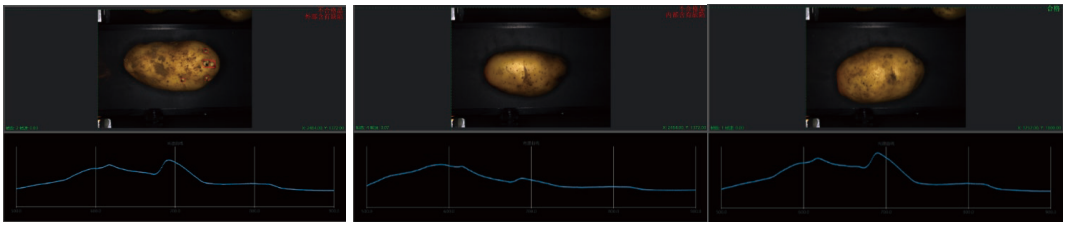
The software system provides redevelopment of the spectrometer and the industrial camera to obtain the image and spectral information, which is then displayed in the software interface. The algorithm predicts external and internal defects in potatoes, with the location of predicted external defects marked using a red box on the potato's image. For internal defects, the algorithm's judgment is verified through potato cutting after flow line detection. Representative results of potato inspection are shown in Figure 11, depicting images with external defects (a), internal defects (b), and no defects (c). The external defect detection algorithm model takes approximately 300 ms, and the internal

defect detection algorithm takes about 50 ms, both of which meet the requirements for online operation. The system achieves an online defect detection rate of approximately 91.3% and can effectively process around 1200 potatoes per hour.



a. Spectral data acquisition part b. Image data acquisition part c. Mechanical part

Figure 10 Hardware System



a. External defects b. Internal defects c. No defects

Note: The upper half of the interface is the image display area. The lower half is the spectrum display area. External defects are selected with red boxes, and the judgment results are displayed in the upper right corner of the interface.

Figure 11 Result of potato detection

4 Conclusions

This study addresses the existing challenges in potato detection in the market by leveraging the YOLOv5s object detection model for detecting external defects such as green skin and rot, and employing a CNN algorithm for detecting internal defects like blackheart disease in potatoes. A potato quality online rapid nondestructive detection system is developed based on the method, and the effectiveness of the method is verified on a rapid flow line, with a defect detection rate of 91.3%. This study is more representative of real-world production scenarios compared to static data collection in the laboratory. Furthermore, this nondestructive detection system, combined with a high-accuracy model, can be quickly applied for other crops as well as for potatoes.

In actual detection scenarios, the missed rate is more critical for evaluating the system performance. Although this study has not extensively discussed strategies for reducing the missed rate. it will be a focal point of future research. The following improvements will be made based on the existing system:

1) Design the feeding device to realize the function of automatic potato feeding;

2) Increase the number of cameras to allow whole surface detection of potatoes;

3) Expand the dataset, collect more samples for model training, and improve the detection accuracy;

4) Conduct further research work on the mechanism.

Acknowledgement

The authors would like to thank all the members of the Research Group for the valuable discussions about the ideas and technical details presented in this paper. This work was supported by the Zhejiang Province Key Research and Development Program (Grant No. 2021C02011), Zhejiang Province Public Welfare Technology Application Research Project (Grant No. LGN18-F030002), Hangzhou Science and Technology Bureau (Grant No. 20201203B116), Program of "Xinmiao" (Potential) Talents in Zhejiang Province (Grant Number: 2022R4-07B055), the Graduate Scientific Research Foundation of Hangzhou Dianzi University (Grant No. CXJJ2022177) and the Major Science and Technology Projects of Breeding New Varieties of Agriculture in Zhejiang Province (Grant No. 2021C02074).

[References]

[1] Xie C H. Potato industry: Status and development. Journal of Huazhong Agricultural University (Social Sciences Edition), 2012; 97(1): 1–4. (in Chinese)

[2] Jing J, Li J W, Liao G P, Yu X J, Viray C. Methodology for potatoes defects detection with computer vision. Proceedings. The 2009 International Symposium on Information Processing (ISIP 2009), 2009; 346p.

[3] Ebrahimi E, Mollazade K, Arefi A. Detection of greening in potatoes using image processing techniques. Journal of American Science, 2011; 7(3): 243–247.

[4] Barnes M, Duckett T, Cielniak G, Stroud G, Harper G. Visual detection of blemishes in potatoes using minimalist boosted classifiers. Journal of Food Engineering, 2010; 98(3): 339–346.

[5] Elmasry G, Cubero S, Moltó E, Blasco J L. In-line sorting of irregular potatoes by using automated computer-based machine vision system. Journal of Food Engineering, 2012; 112(1-2): 60–68.

[6] Oppenheim D, Shani G. Potato disease classification using convolution neural networks. Advances in Animal Biosciences, 2017; 8(2): 244–249.

[7] Elsharif A A, Dheir I M, Mettleq A, Abu-Naser S S. Potato classification using deep learning. International Journal of Academic Pedagogical Research, 2020; 3(12): 1–8.

[8] Chen J D, Zhang D F, Nanehkaran Y A, Li D L. Detection of rice plant diseases based on deep transfer learning. Journal of the Science of Food and Agriculture, 2020; 100(7): 3246–3256.

[9] Zhao G Y, Quan L X, Li H L, Feng H Q, Li S W, Zhang S H, et al. Real-time recognition system of soybean seed full-surface defects based on deep learning. Computers and Electronics in Agriculture, 2021; 187: 106230.

[10] Ramos R P, Gomes J S, Prates R M, Filho E F, Teruel B J, Costa D S. Non-invasive setup for grape maturation classification using deep learning. Journal of the Science of Food and Agriculture, 2021; 101(5): 2042–2051.

[11] Thuyet D Q, Kobayashi Y C, Matsuo M. A robot system equipped with deep convolutional neural network for autonomous grading and sorting of

root-trimmed garlics. Computers and Electronics in Agriculture, 2020; 178: 105727.

[12] Nouri-Ahmadabadi H, Omid M, Mohtasebi S S, Firouz M S. Design, development and evaluation of an online grading system for peeled pistachios equipped with machine vision technology and support vector machine. Information Processing in Agriculture, 2017; 4(4): 333–341.

[13] Blasco J, Cubero S, Gómez-Sanchís J, Mira P, Moltó E. Development of a machine for the automatic sorting of pomegranate (*Punica granatum*) arils based on computer vision. Journal of Food Engineering, 2009; 90(1): 27–34.

[14] Hajjar G, Quellec S, Pépin J, Challois S, Joly G, Deleu C, et al. MRI investigation of internal defects in potato tubers with particular attention to rust spots induced by water stress. Postharvest Biology and Technology, 2021; 180: 111600.

[15] Sosa P, Guild G, Burgos G, Bonierbale M, Felde T. Potential and application of X-ray fluorescence spectrometry to estimate iron and zinc concentration in potato tubers. Journal of Food Composition and Analysis, 2018; 70: 22–27.

[16] López-Maestresalas A, Keresztes J C, Goodarzi M, Arazuri S, Jarén C, Saeys W. Non-destructive detection of blackspot in potatoes by Vis-NIR and SWIR hyperspectral imaging. Food Control, 2016; 70: 229–241.

[17] Wu L G, He J G, Liu G S, Wang S L, He X G. Detection of common defects on jujube using Vis-NIR and NIR hyperspectral imaging. Postharvest Biology and Technology, 2016; 112: 134–142.

[18] Ye D D, Sun L J, Tan W Y, Che W K, Yang M C. Detecting and classifying minor bruised potato based on hyperspectral imaging. Chemometrics and Intelligent Laboratory Systems, 2018; 177: 129–139.

[19] Jamshidi B, Minaei S, Mohajerani E, Ghassemian H. Reflectance Vis/NIR spectroscopy for nondestructive taste characterization of Valencia oranges. Computers and Electronics in Agriculture, 2012; 85: 64–69.

[20] Peirs A, Lammertyn J, Ooms K, Nicolaï B M. Prediction of the optimal picking date of different apple cultivars by means of VIS/NIR-spectroscopy. Postharvest Biology and Technology, 2001; 21(2): 189–199.

[21] Tušek A J, Benković M, Malešić E, Marić L, Jurina T, Kljusurić J G. Rapid quantification of dissolved solids and bioactives in dried root vegetable extracts using near infrared spectroscopy. Spectrochimica acta Part A: Molecular and Biomolecular Spectroscopy, 2021; 261: 120074.

[22] Moomkesh S, Mireei S A, Sadeghi M, Nazeri M. Early detection of freezing damage in sweet lemons using Vis/SWNIR spectroscopy. Biosystems Engineering, 2017; 164: 157–170.

[23] Li J B, Huang W Q, Zhao C J, Zhang B H. A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy. Journal of Food Engineering, 2013; 116(2): 324–332.

[24] Scalisi A, O'Connell M G. Application of Visible/NIR spectroscopy for the estimation of soluble solids, dry matter and flesh firmness in stone fruits. Journal of the Science of Food and Agriculture, 2021; 101(5): 2100–2107.

[25] Wang F, Li Y Y, Peng Y, Yang B N, Li L, Yin X Q. Hand-held device for non-destructive detection of potato quality parameters. Transactions of the CSAM, 2018; 49(7): 348–354.

[26] Wang F, Li Y Y, Peng Y K, Yang B N, Li L, Liu Y C. Multi-parameter potato quality non-destructive rapid detection by visible/near-infrared spectra. Spectroscopy and Spectral Analysis, 2018; 38(12): 3736–3742.

[27] Zhang X Y, Liu W, Xing L, Zhao F M, Yang Y C, Yang B N. An near-infrared prediction model for quality indexes of potato processing. Infrared, 2012; 33(12): 33–39.

[28] Zhu Z, Zeng S W, Li X Y, Zheng J. Nondestructive detection of blackheart in potato by visible/near infrared Transmittance Spectroscopy. Journal of Spectroscopy, 2015; 2015: 1–9.

[29] Song Y, Wang X Z, Xie H L, Li L Q, Ning J M, Zhang Z Z. Quality Evaluation of Keemun black tea by fusing data obtained from near-infrared reflectance spectroscopy and computer vision sensors. Spectrochimica Acta Part A:Molecular and Biomolecular Spectroscopy, 2021; 252(5): 119522.

[30] Yu H L, Liang Y L, Liang H, Zhang Y Z. Recognition of wood surface defects with near infrared spectroscopy and machine vision. Journal of Forestry Research, 2019; 30(6): 2379–2386.

[31] Huang X Y, Xu H X, Wu L, Dai H, Yao L Y, Han F K. A data fusion detection method for fish freshness based on computer vision and near-infrared spectroscopy. Analytical Methods, 2016; 8(14): 2929–2935.

[32] Yin J F, Hameed S, Xie L J, Ying Y B. Non-destructive detection of foreign contaminants in toast bread with near infrared spectroscopy and computer vision techniques. Journal of Food Measurement and Characterization, 2021; 15(1): 189–198.

[33] Ministry of Agriculture of the PRC. Grades and specifications of potatoes, 2006. Available: http://www.moa.gov.cn/govpublic/SCYJJXXS/201006/t20100606_1532967.htm.

[34] Han Y F, Lü C X, Yuan Y W, Yang B N, Zhao Q L, Cao Y F, et al. PLS-discriminant analysis on patato blackheart disease based on VIS-NIR transmission spectroscopy. Spectroscopy and Spectral Analysis, 2021; 41(4): 1213–1219.

[35] Jubayer F, Soeb J A, Mojumder A N, Paul M K, Barua P, Kayshar S, Akter S S, et al. Detection of mold on the food surface using yolov5. Current Research in Food Science, 2021; 4: 724–728.

[36] Rong D, Wang H Y, Ying Y B, Zhang Z Y, Zhang Y S. Peach variety detection using VIS-NIR spectroscopy and deep learning. Computers and Electronics in Agriculture, 2020; 175: 105553.

[37] Tazehkandi A A. Computer Vision with OpenCV 3 and Qt5: Build visually appealing, multithreaded, cross-platform computer vision applications. Packt Publishing Ltd, 2018; 486p.

[38] Yao L J, Lu L, Zheng R. Study on detection method of external defects of potato image in visible light environment. 2017 10th International Conference on Intelligent Computation Technology and Automation (ICICTA), 2017; pp.118–122.

[39] Clark C J, McGlone V A, Jordan R B. Detection of brownheart in 'braeburn' apple by transmission NIR spectroscopy. Postharvest Biology and Technology, 2003; 28(1): 87–96.