AUTOMATED GRADING SYSTEM FOR SWEET POTATO INDUSTRY USING APRILTAG AND EDANET

Ko-Hua Chen[1]

[1]Graduate Student, Department Civil Engineering, National Chung Hsing University
Rm. 7, 3F., No. 145, Xingda Rd., South Dist., Taichung City, 402, Taiwan
Email: g111062402@mail.nchu.edu.tw

Ming-Der, Yang[2*]

[2*] Professor, National Chung Hsing University
Rm. 7, 3F., No. 145, Xingda Rd., South Dist., Taichung City, 402, Taiwan
Email: mdyang@dragon.nchu.edu.tw

KEY WORDS: computer vision, sweet potato, grading system

ABSTRACT: As the trend of population aging and declining labor force becomes increasingly significant, according to observations in the sweet potato industry, the sweet potato industry in Taiwan is dominated by middle-aged labor-intensive and long working hours. Among them, sweet potato grading is mainly done through labor selection, and there is no shortage of misclassification due to fatigue and human error.

The production end needs to respond to these challenges and improve production efficiency. In this regard, automatic object recognition and classification through computer vision grading systems has become an important solution in the industry in recent years.

This study proposes a method based on AprilTag combined with the segmentation model EDANet, which uses EDANet to accurately segment the object position and calculates the scale in the real image by recognizing the position, so as to calibrate the image and obtain accurate object size, providing a solution to this problem. AprilTag is a special marker that can be easily detected and identified in images. This study uses the position information of AprilTag in the image, combined with the corresponding scale conversion, to achieve accurate calculation of object size in the image.

Through experiments and tests, it is proved that the accuracy of the method proposed in this study. Using AprilTag for position recognition and scale conversion can accurately obtain object size information, thereby achieving more accurate object grading and production management. This has important significance for responding to challenges such as population aging and declining labor force, improving production efficiency and management level.

In summary, this study proposes a method based on AprilTag combined with the segmentation model EDANet, which calculates the scale in the real image by recognizing the position of AprilTag, so as to calibrate the image and obtain the accurate size of the required object. The application and development of this computer vision grading system under the background of population aging and declining labor force can have a significant impact.

1. INTRUCTION

Sweet potato (Ipomoea batatas), ranked as the seventh-largest staple crop globally, boasts a cultivation area of approximately 9 million hectares. Its popularity is attributed to its ease of cultivation and high nutritional content. The entire sweet potato plant is edible, with the storage root being the primary consumable part. The storage root primarily comprises carbohydrates, dietary fiber, and essential minerals such as potassium, calcium, and vitamin C. The journey of sweet potatoes from the farm to consumers' tables involves several processing steps, including initial selection at the production site, machine cleaning, preliminary sorting, manual sorting, and final distribution to meet various consumer needs. Storage roots with defects, often due to food industry demands, are sorted out during this process.

Specifically, sweet potatoes are harvested from the soil, cleaned, preliminarily screened, graded, and processed according to different requirements, such as storage or peeling, before reaching consumers. During the initial screening at processing plants, a significant amount of manual labor is involved in picking out damaged sweet potatoes caused by transportation damage, rodent bites, insect damage, or whitefly disease infection. These sweet potatoes are transported on conveyors at speeds of approximately 0.1 to 0.2 m/s. The selection of damaged sweet potatoes primarily relies on human visual recognition, which can impact quality control and result in defective sweet potatoes reaching customers.

Additionally, processing plants categorize sweet potatoes based on weight according to customer demands. Existing sorting machines follow the standard for potatoes, allowing sorting only based on fixed diameters, leading to manual sorting in most cases for sweet potatoes. In contrast, the potato industry, a similar sector, has embraced mature image recognition technology using computer vision to accelerate defect detection and sorting processes, reducing the chances of oversight-related omissions. Moreover, labor involved in the screening process often require training to acquire the necessary inspection skills. New employees may not immediately acquire the inspection abilities established by their predecessors, and different individuals may have varying screening standards, potentially resulting in suboptimal screening effectiveness. Therefore, establishing a standardized screening algorithm is essential not only for uniform quality inspection but also for enhancing production efficiency.

In the context of sweet potato image recognition, as there are no well-established relevant applications for grading the sweet potatoes, this study aims to establish an image recognition technology based on semantic segmentation neural network models for grading purposes. Aim to solve:

1. Accelerate the recognition and grading process to enhance overall production yield.
2. Liberate labor resources to enable more efficient workforce allocation.
3. Establish a standardized recognition criterion.
4. Establish a comprehensive workflow for subsequent defect segmentation.

The primary contributions of this study are as follows:

1. Application of a neural network based on semantic segmentation with asymmetric convolutions for sweet potato grading recognition, followed by performance evaluation.
2. Establishment of a comprehensive workflow for sweet potato grading detection on the production line.
3. Taiwan's first sweet potato annotated image dataset, comprising 187 images containing sweet potatoes of various sizes.

2. LITERATURE REVIEW

With the advancement of computer vision and image processing in recent years, many industries that previously relied heavily on manual labor have found liberation through these methods [9]. Various algorithms related to computer vision have been applied in grading and fruit selection across different sectors.

Dang et al. 2010, employed CMOS cameras for edge detection and size estimation of apples. Bargoti et al. 2016, utilized multi-scale multi-layer perceptrons (ms-MLP) and convolutional neural networks (CNN) for segmentation and yield estimation of apple orchard images, achieving an F1 score of 0.791.TheOo et al. 2020, estimated mango volume by identifying the long axis position in HSV mango images, slicing it width-wise, and calculating the estimated volume using the ellipsoid volume formula, with an average accuracy of 96.8%.Panprasittikit et al. ,2021, classified citrus sizes based on binarized citrus images using machine learning and neural networks, achieving the highest accuracy of 90.71% with support vector machines (SVM). Korchagin et al. 2021, employed a two-stage algorithm, first using the Viola-Jones algorithm to filter camera images and then using machine learning and deep learning methods (SIFT-SVM, HOG-CNN, Otsu’s Threshold Binarization—CNN) to detect defects. The results demonstrated accuracy ranging from 80% to 97%, with the ability to simultaneously detect up to 100 tubers. Dolata et al. 2021, applied the Mask R-CNN instance segmentation model to segment potatoes, achieving a panoramic quality measure (PQ) of 0.889. They further obtained a regression model with an R-squared score of 0.869 for diameter estimation from images.

In summary, research in image processing, deep learning, and machine learning has been effectively applied in the context of vegetables and fruits. However, there is limited research specifically focused on sweet potatoes. The present study addresses this gap and holds potential for making significant contributions in this domain.

3. MATERIAL AND METHOD

Type text single-spaced, with one blank line between paragraphs and following headings. Start paragraphs flush with left margin.

3.1 Material

The imaging system used in this study was constructed using corrugated cardboard to create a dark box, as depicted in Figure 1. The primary parameters of the image acquisition equipment are detailed in Table 1. An RGB camera from an iPhone 7 Plus was utilized, with the following camera settings: Shutter Speed: 1/643s, Aperture: f/1.8, ISO: 100. The camera lens was oriented vertically downward at a distance of 450 millimeters from the sweet potatoes. LED light bars provided the illumination, directing light vertically onto the surface of the sweet potatoes, with a color temperature of 4500K. To minimize light reflections, the interior of the box was lined with light-blocking fabric. The purpose of installing the light source was to supply illumination within the dark box and ensure that external lighting did not interfere.

Table1.Capture parameters

Camera	iPhone 7 Plus
Shutter Speed	1/643s
Aperture	f/1.8
ISO	100
Distance	450mm
color temperature	4500K

Figure 1.DIY Chamber

To simulate the future effects of industrial-grade cameras for capturing and segmenting sweet potato images, mobile phones with similar-sized image sensors were used for image capture. Each sweet potato was individually weighed using a scale, and the weights were recorded. A total of 187 images were collected. Following the manufacturer's classification standards, the selected grading primarily consisted of three categories based on weight, Large Grade: 90g and above (63 images)- Medium Grade: 80g~90g (63 images)- Small Grade: 80g and below (60 images).

3.2 Methods

To obtain the real-world scale for AprilTags in imaging environment, a specific tag, "tag25h9," with a known side length of 10cm is used. This tag was captured in the same photography setup as the sweet potatoes, with identical camera parameters. Using a Python API package, and extracted the pixel coordinates of the four tag corners in the image. By dividing the known real-world tag side length by the measured pixel side length, determining a scaling factor. This scaling factor represents the ratio of pixels to actual distances in that specific environment.

For actual recognition, as shown in Figure 2, sweet potatoes were segmented in the image under the same photography conditions. The segmented sweet potatoes' pixel areas were calculated and multiplied by the scaling factor obtained from the tag, yielding the actual area in the image. This real area information was then used in conjunction with regression results to classify the sweet potatoes.

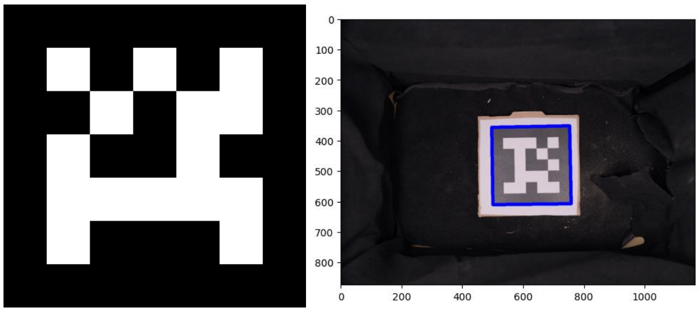


Figure 2 tag25h9 and the recognition screen

3.3 Segmentation model

With the recent advancements in deep learning, convolutional neural network (CNN) models have demonstrated significant capabilities in image recognition tasks. In this study, we adopted the EDANet model based on deep convolutional neural networks (as shown in Figure 3). This model was chosen due to its distinctive features, including the use of asymmetric convolutions, dilated convolutions, and dense connections, which enhance both computational speed and predictive capabilities. Additionally, considering future deployment on edge computing devices within production lines, a model with a low parameter count is compatible with devices that may not have high-capacity memory. This model contains only 680,000 parameters and can achieve a processing speed of 108fps on Nvidia GPUs 1080 when inferring images of size 512x1024 from the Cityscapes dataset. Taking into account computational power and parameter count, EDANet stands out as the optimal choice.

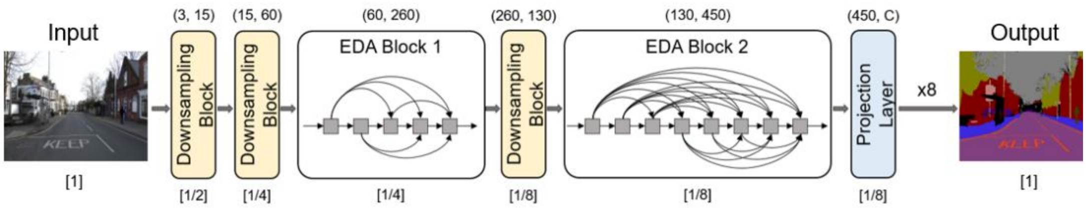


Figure 2: The proposed EDANet architecture. The numbers of input and output channels of each block are marked in parenthese The numbers in brackets are output feature size ratios to the full-resolution input images. “C”: the number of object classes.

Figure 3. EDANet architecture

3.4 Acknowledgements (optional)

The training dataset consists of a total of 187 images containing sweet potatoes (Figure 4). Half of the images were used for training and validation, with 80% of this dataset used for training images and 20% for validation images. The other half of the dataset was used as the testing dataset. Image annotation was performed using the GIMP software, an open-source raster graphics editor. Background and sweet potatoes were labeled in blue (RGB value: [0,0,255]) and green (RGB value: [0,255,0]), respectively. To reduce image size without affecting image resolution, the captured images were initially cropped to include a 1500x1400 region containing the sweet potatoes. During training of the semantic segmentation model, these images were further scaled down to 512x512 to improve computational speed and increase batch size. During the inference phase, the original cropped images of 1500x1400 were used to preserve the original image resolution.

The training equipment used was a desktop computer with an Intel(R) Xeon(R) W-2235 CPU @ 3.80GHz and an Nvidia RTX A4000 graphics accelerator card with 16GB of GPU memory. The framework utilized was TensorFlow, developed by Google, with version 2.10. CUDA version 11.6 was used for GPU acceleration during training. The optimizer employed was the Adam optimizer with a learning rate of 0.005 and a batch size of 10. After training, the model achieved an overall accuracy of 0.99, precision of 0.92, recall of 1.0, and an F1 score of 0.96 on the testing dataset.

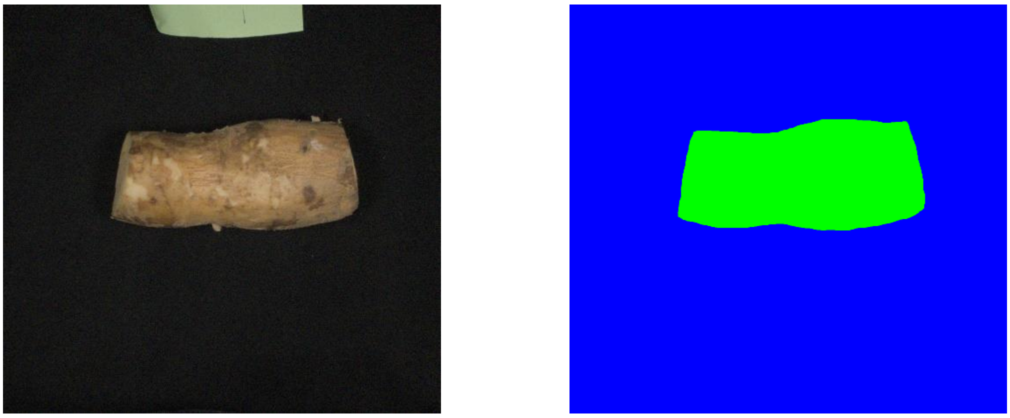


Figure 3: Training and annotation image of size 1500x1400

4. EXPERIMENT

The training data, after training the segmentation model, was segmented into sweet potato images using EDANet. The segmentation results were used to build regression models for sweet potato weight estimation. Four regression models were employed, including Linear Regression, XGBoost, Random Forest Regression, and K-Nearest Neighbors (K-NN) Regression. For each model, parameter exploration was conducted to determine the optimal settings based on the highest R-squared value. The following parameters were selected for the best R^2 values: Linear Regression: Degree is set to 19, resulting in R^2 values of 0.838 (degree 1) and 0.841 (degree 19). XGBoost Algorithm: Estimators is set to 32, Learning Rate :0.1, with an R^2 value of 0.736. Random Forest Regression's n_estimators:66, yielding an R^2 value of 0.769. K-Nearest Neighbors (K-NN) algorithm's neighbors is set to 7, resulting in an R^2 value of 0.824. Support Vector Machine (SVM) classifier set to 3, with an R^2 value of 0.772. After building the regression models and estimating the weights, sweet potatoes were classified into different grades based on the standards: Large grade (≥ 90g), Medium grade (80g~90g), and small grade (< 80g). The accuracy of each regression model for the final classification was then calculated.

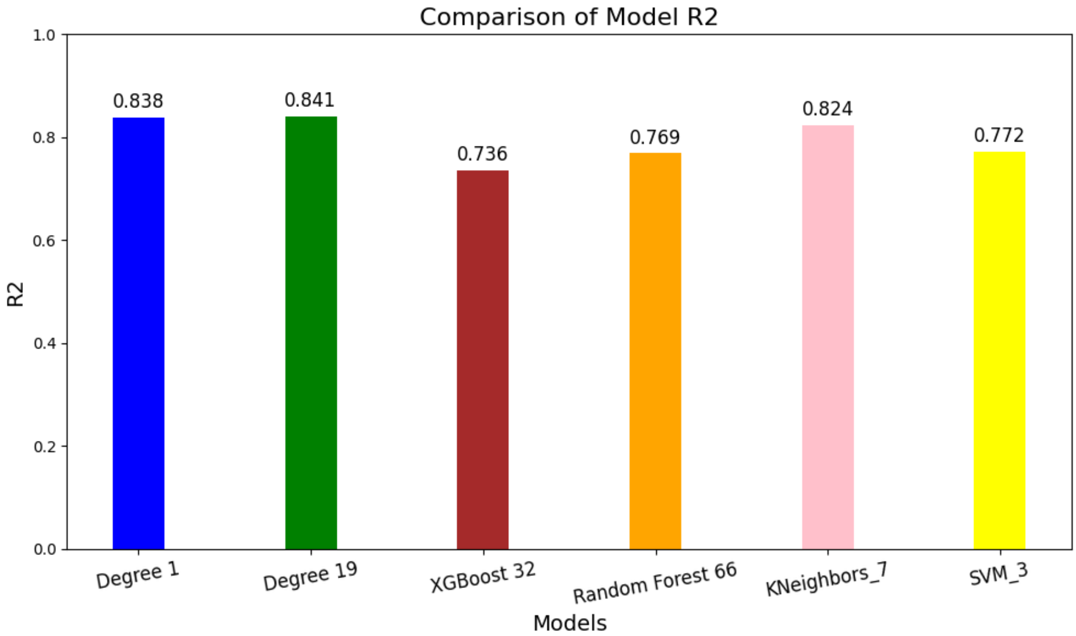


Figure 4: R-squared Scores of Regression Models.

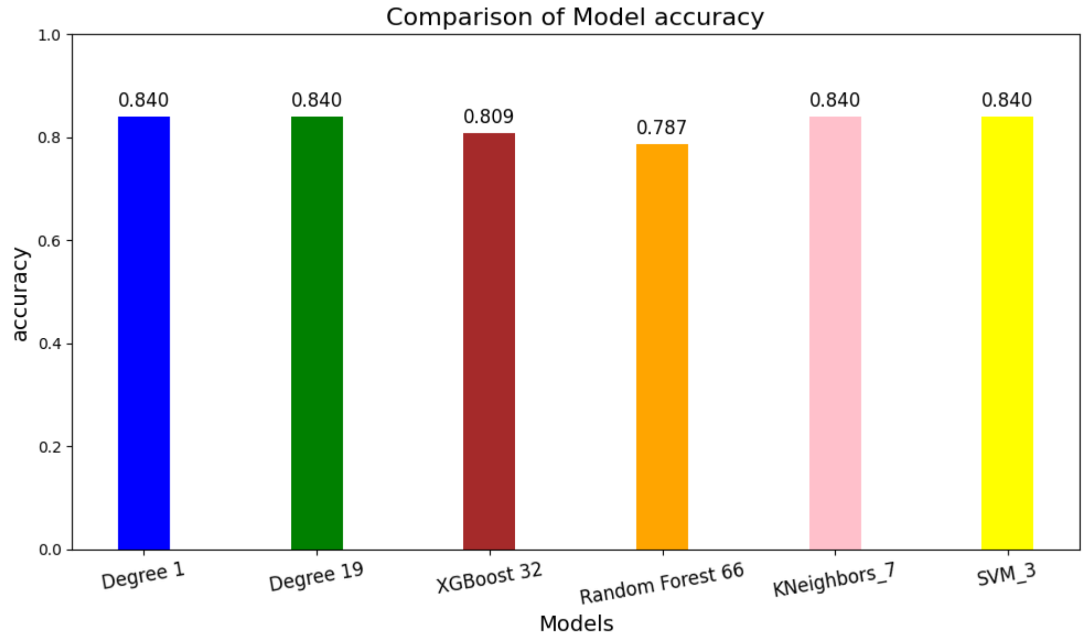
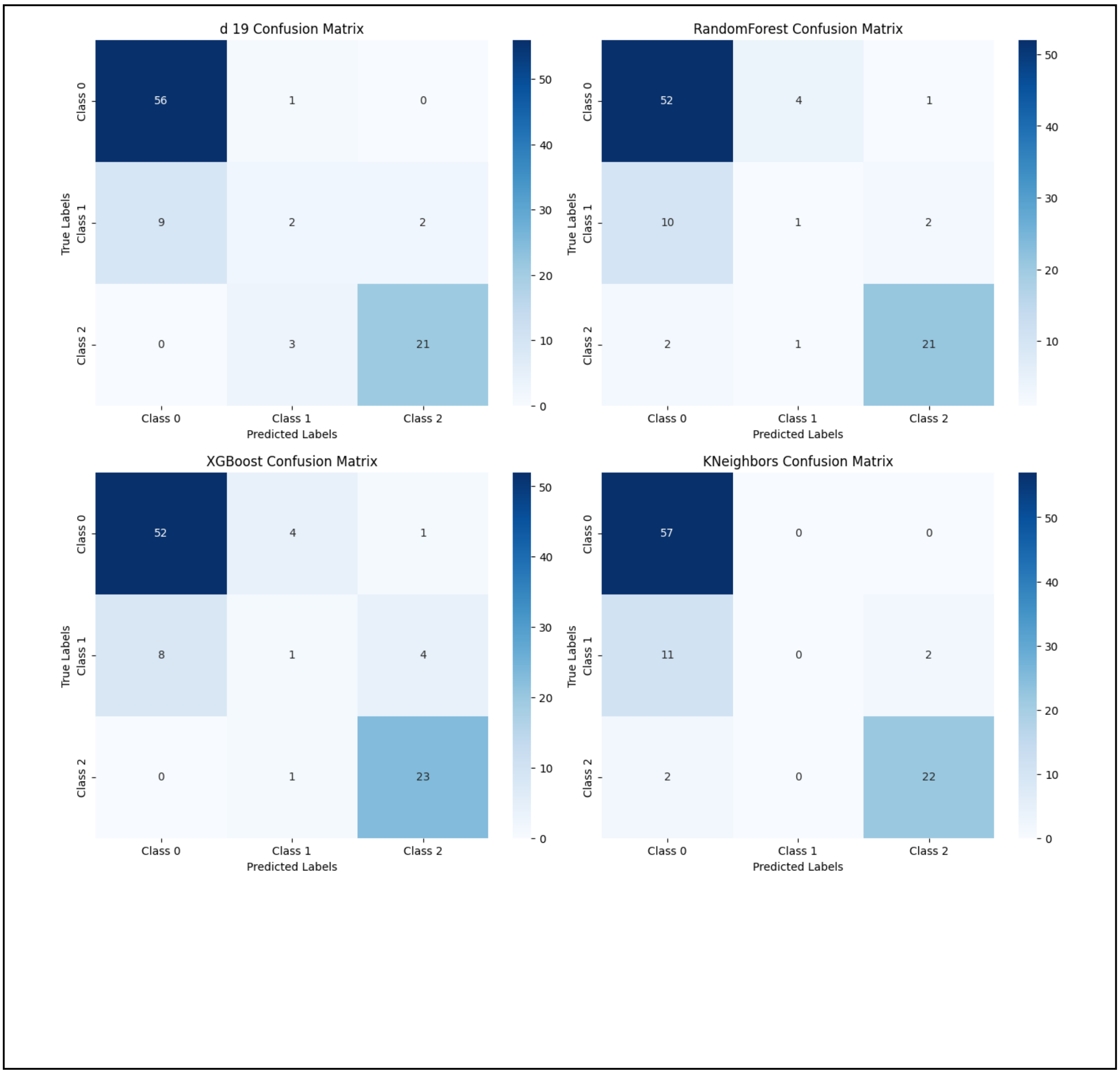


Figure 5: Classification Accuracy of Regression Models.



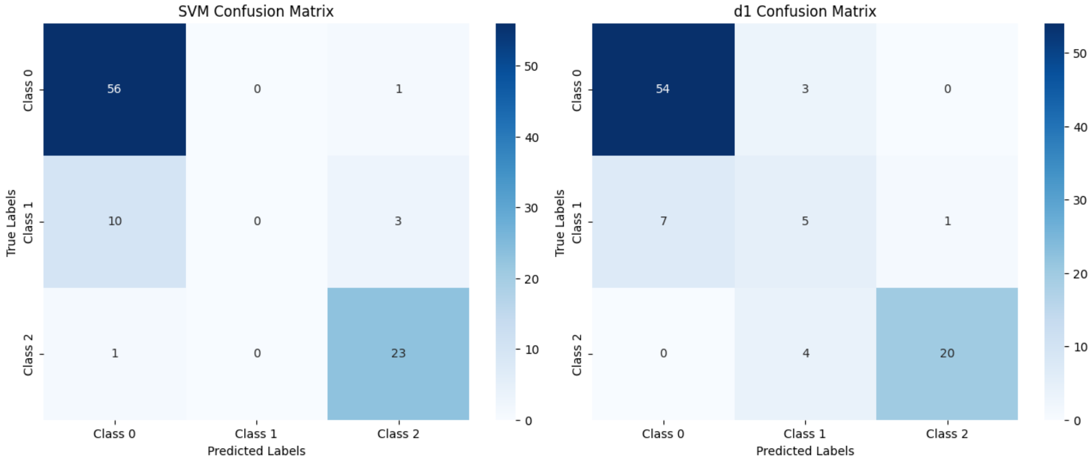


Figure 6: Confusion Matrices of Regression Models for Each Category.

5. CONCLUSION

The study proposes a classification algorithm for sweet potato weight on the production line, which uses regression models based on semantic segmentation results to estimate sweet potato weights and classify them according to different weight standards. Overall, the sweet potato classification and identification method based on semantic segmentation neural networks hold great promise for widespread practical applications in the sweet potato industry, providing a crucial tool for production optimization.

Furthermore, future research can extend the application of this classification and identification model to sweet potato defect recognition. By integrating image processing techniques and machine learning algorithms, accurate detection and classification of defects, damages, and diseases on the surface of sweet potatoes can be achieved. This would enable the simultaneous selection and classification of sweet potato defects during the production process, offering a more comprehensive and efficient production solution for the sweet potato industry. Such an integrated model not only helps reduce human error and labor intensity in manual selection but also enhances the overall quality and competitiveness of the products. Through further research and development, this comprehensive model has the potential to bring intelligence and automation to agricultural production, injecting new vitality into the sustainable development of the sweet potato industry.

References

Bargoti, Suchet and James Patrick Underwood. 2016, Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards. *Journal of Field Robotics* 34 n. pag.

E. Olson, 2011,AprilTag: A robust and flexible visual fiducial system, 2011, IEEE International Conference on Robotics and Automation, Shanghai, China, pp. 3400-3407,

H. Dang, J. Song and Q. Guo, 2010, A Fruit Size Detecting and Grading System Based on Image Processing, Second International Conference on Intelligent Human-Machine Systems and Cybernetics, Nanjing, China, 2010, pp. 83-86,

J. Panprasittikit and K. Amnatchotiphan, 2021, Thai Tangerine Size Classification via Computer Vision, 2021 IEEE 8th International Conference on Industrial Engineering and Applications (ICIEA), Chengdu, China, pp. 571-574

Korchagin, S.A.; Gataullin, S.T.; Osipov, A.V.; Smirnov, M.V.; Suvorov, S.V.; Serdechnyi, D.V.; Bublikov, K.V. 2021, Development of an Optimal Algorithm for Detecting Damaged and Diseased Potato Tubers Moving along a Conveyor Belt Using Computer Vision Systems. *Agronomy*, 11, 1980

Przemysław Dolata, Paweł Wróblewski, Mariusz Mrzygłód, Jacek Reiner, 2021, Instance segmentation of root crops and simulation-based learning to estimate their physical dimensions for on-line machine vision yield monitoring, *Computers and Electronics in Agriculture*, Volume 190

Shao-Yuan Lo, Hsueh-Ming Hang, Sheng-Wei Chan, and Jing-Jhih Lin. 2020, Efficient Dense Modules of Asymmetric Convolution for Real-Time Semantic Segmentation. In Proceedings of the ACM Multimedia Asia (MMAsia '19). Association for Computing Machinery, New York, NY, USA, Article 1, 1–6.

TheOo Mon, Nay ZarAung, 2020, Vision based volume estimation method for automatic mango grading system, *Biosystems Engineering*, Volume 198, Pages 338-349