Development of a machine vision system to estimate the physical attributes of potato tubers on-the-go at the post-harvest stage

by

Ighodaro K. Emwinghare

Submitted in partial fulfilment of the requirements

for the degree of Master of Science

at

Dalhousie University

Halifax, Nova Scotia

July 2023

Table of Contents

List of Tables

List of Figures

Abstract

This study presents a deep learning and image processing-based machine vision system for sampling and sizing full-size potato tubers on post-harvest conveyors. First, we present a method for sampling fully visible potato tubers running on post-harvest conveyors in the Laboratory and the field, overcoming challenges such as occlusion and varying lighting conditions. This method utilizes Mask R-CNN and image feature-based machine learning models, achieving high sampling accuracy and segmentation quality that averaged over 90% even in field conditions. Subsequently, a machine vision system designed to estimate the size of potato tubers sampled on the post-harvest conveyor is proposed. To validate the efficacy of this proposed system, two distinct methods were employed: static and dynamic conveyor experiments. The outcomes of these experiments revealed a minimum coefficient of determination of 0.77 for the estimation of the minor diameter of the potato tubers when they were in free-rolling motion on the conveyors, regardless of their orientation and spatial arrangement within clusters. Furthermore, the dimension errors observed across all scenarios remained consistently below 10%, affirming the system's accuracy and robustness in size estimation.

List of Abbreviations

AR	Average Recall
AP	Average Precision
AUROC	Area Under the Receiver Operating Characteristic Curve
CCC	Lin's Concordance Correlation
CCD	Charged-Couple Device
CNN	Convolutional Neural Network
COCO	Common Objects in Context
CT	Cathode-ray Tube
Deep SORT	Deep Simple Online Realtime Tracking
DNN	Deep Neural Network
FAO	The Food and Agriculture Organization
Faster R-CNN	Faster Region-Based Convolutional Neural Network
FCN	Fully Convolution Network
FPN	Feature Pyramid Network
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
HSV	Hue Saturation Value
KDE	Kernel Density Estimate
LDA	Linear Discriminant Analysis
LR	Logistic Regression
LSTM	Long Short-term Memory Networks
L1	Ridge Regression
L2	Lasso Regression
Mask R-CNN	Mask Region-Based Convolutional Neural Network
mIoU	Mean Intersection over Union
MLF-NN	Multilayer Feed Forward Neural Network
Mlxtend	Machine Learning Extensions

OpenCV	Open Source Computer Vision Library
RBF	Radial Basis Function
RCNN	Region-Based Convolutional Neural Network
ResNet	Residual Network
RF	Random Forest
R-FCN	Region-based Fully Convolutional Networks
RGB	Red Green Blue
RGB-D	Red Green Blue-Depth
RMSE	Root Mean Square Error
nRMSE	Normalized Root Mean Square Error
ROI	Region of Interest
RPN	Region Proposal Network
R^2	Coefficient of Determination
SFSS	Sequential Forward Feature Selection
SSD	Single-Shot Detector
SVM	Support Vector Machine
S/N_1	Signal-to-Noise ratio 1
S/N_2	Signal-to-Noise ratio 2
UV	Ultraviolet
VGG	Visual Geometry Group
YOLO	You Only Look Once
YOLOv5	You Only Look Once version 5
3D	3 Dimensional

Acknowledgments

I want to express my sincere gratitude to Dr. Ahmad Al-Mallahi, my supervisor, for his unwavering support and guidance throughout my master's program. Dr. Al-Mallahi has been readily available to address my numerous inquiries and has provided me with the necessary direction to ensure my success. Additionally, I would like to acknowledge the valuable contributions of Dr. Travis Esau and Dr. Felipe Campelo, members of my supervisory committee, who offered insightful input and graciously accommodated my requests despite their busy schedules.

I am deeply grateful to the employees at McCain Foods, specifically Manphool Fageria and Yves Leclerc, for granting me access to the resources at the McCain Farms of the Future in Florenceville-Bristol. Their provision of essential information has been instrumental in comprehending the industry relevance of my research.

Furthermore, I extend my heartfelt appreciation to my colleagues in the Applied Intelligence Engineering Systems Research Laboratory. Reem Abukmeil, Imran Hassan, Mozammel Motalab, Sama Huseynova, Colton Campbell, and Humphrey Maambo have provided invaluable assistance at various stages of my research, for which I am truly grateful. I also want to thank my housemates, Taiwo Makinde and Taiwo Erinle, for inspiring me, especially in their pursuit of academic excellence.

Chapter 1: Introduction

1.1 Background

Potatoes (Solanum tuberosum) are graded commercially based on their physical attributes and chemical composition. This grading process entails assessing several characteristics, such as shape, size, surface, and interior defects, as well as the sugar and starch content of the tubers (Sanchez et al., 2020). These quality features are critical in determining the market value and potential uses of the potatoes, as different grades are used for specific purposes (Rady & Guyer, 2015). The physical characteristics of potato tubers, including their size and shape, influence their intended applications. Large and elongated tubers are valued for producing French fries and chips due to their uniform dimensions, which contribute to creating visually appealing end products (Kabira & Lemaga, 2003). Conversely, smaller tubers with surface defects or irregular shapes are used for other processed products, such as mashed potatoes or potato flakes (Abong et al., 2009; Marwaha et al., 2010). Therefore, grading based on physical attributes is critical for efficiently classifying and allocating potatoes to their respective market segments.

The timely grading of potato tubers following harvest is essential for several reasons. Firstly, it allows setting appropriate prices based on crop quality, ensuring farmers are fairly compensated for their produce. Secondly, precise grading allows for the optimal distribution of potatoes to various processing factories or market channels, reducing waste and increasing crop utilisation. Finally, accurate grading of potato tubers provides valuable information for farmers and processors, allowing them to plan and organize their logistical operations efficiently. By having early knowledge of the quality attributes of the harvested crop, stakeholders can make well-informed decisions regarding storage, transportation, and processing requirements. This information-driven approach enhances operational

efficiency and helps mitigate post-harvest losses, ultimately ensuring that potatoes reach consumers in optimal condition.

In the province of New Brunswick in Canada, potatoes are harvested for four weeks during the first two months of the Fall season to avoid damage from frost. To maximize harvest efficiency, the tubers are transported into storage using conveyors, with a throughput of approximately 545 kg per minute, thus, leading to clustering and occlusion of the tubers on the post-harvest conveyors. Moreover, the field environment introduces additional factors such as mechanical vibration, ambient light, and the presence of foreign objects, which can vary across different sections of the farmland, including at the harvester. As a result, grading potato tubers in the field is a complex task. Furthermore, grading every tuber harvested would be time-consuming and resource-intensive, requiring a significant workforce and a vast amount of time. Therefore, to simplify the grading process and make it more manageable, a representative sample of tubers is selected for the size grading of the crop (Pavlista & Ojala, 1997).

Based on observations of post-harvest operations at the Farms of the Future of McCain Foods in Florenceville-Bristol, New Brunswick, a selection of approximately 35 kg of potato tubers per 175 tonnes is graded for quality. The selected tubers are manually singulated and passed one after the other through a Gocator 3D sensor (LMI Technologies, Vancouver, Canada) for size estimation based on laser triangulation and fringe pattern projection and subsequently examined for surface defects such as greening and internal defects such as hollow heart and blackleg. While the Gocator 3D sensor has high accuracy, falling within the range of 0.0018 mm to 0.0030 mm (Xiong et al., 2016), it operates at a fixed conveyor speed and cannot provide precise measurements when the tubers are in

contact with one another. Consequently, manual intervention becomes necessary, leading to a labour-intensive, time-consuming, and cost-intensive process, particularly during the harvest season when the workforce is limited. Moreover, due to the limited sampling frequency and sample size, the tubers selected for grading may not represent the entire range of harvest variability. Given these challenges, there is an evident need to explore alternative solutions that can enhance the assessment of quality attributes of potato tubers.

1.2 Machine Vision

The use of machine vision for quality grading tasks has attracted significant attention due to its affordability, ease of integration, and repeatability (Dolata et al., 2021; Ismail & Malik, 2022; Su et al., 2020). Various imaging technologies, including charge-coupled device (CCD) cameras, hyperspectral cameras, ultra-violet (UV) and X-ray cathode-ray tube (CT) cameras, and other approaches, have been previously deployed to identify critical potato quality features (Su et al., 2020). The physical parameters of potato tubers, such as their length, width, and mass, can already be estimated by machine vision systems (Su et al., 2017). However, implementing these machine vision systems in field conditions without incurring additional costs or requiring significant adjustments to the existing system remains challenging.

Researchers have addressed some of the challenges of grading potato tubers in the field in recent years, particularly those related to complex backgrounds and unstructured lighting conditions. Al-Mallahi et al. (2008) studied the influence of the potato conditions (whether dry or wet) on tuber detection while being harvested using cameras and built a machine vision system to detect potatoes on the harvester (Al-Mallahi et al., 2010b) including the condition under which the potato tubers and clods may be clustered in-line (Al-Mallahi et

al., 2010a). Smith et al. (2018) also explored the use of an RGB-D camera to size-grade potato tubers on a harvester with customized housing, to minimize the effects of direct sunlight and harsh weather conditions. Similarly, Lee & Shin (2020) proposed a machine vision system that utilized conventional image processing techniques for estimating the mass of potato tubers on the field surface after being dug by a potato digger. Despite their promising results, these investigations are customized to specific challenges and rely on manually crafted image features, which may not generalize to the spatial and temporal variations encountered in the field. Moreover, none of the studies addressed the challenge of mutually occluding tubers, often encountered at different sites in the potato field.

The application of deep learning-based machine vision systems in grading tasks has become increasingly popular due to their ability to learn intricate features (Koirala et al., 2019). This capability allows them to address the variations encountered in the field effectively. Dolata et al. (2021) and Lee & Shin (2020) employed Mask R-CNN, a deep learning algorithm for instance segmentation, for yield assessment of potato tubers. Their respective studies aimed to overcome challenges related to the size estimation of tubers on the harvester and on the field surface after excavation by a potato digger. Notably, these studies showed the effectiveness of deep learning algorithms in handling various complexities, including complex backgrounds, clustering, foreign objects, and ambient light. Nevertheless, both studies encountered challenges in detecting potato tubers that were significantly occluded. Moreover, accurately assessing the occluded portion of the tubers is difficult.

Instead of assessing every tuber, regardless of whether they are occluded or not, a pragmatic and economical approach to tackle the occlusion issue could be the development

of a technique that can differentiate fully visible tubers from those that are occluded. By exclusively utilizing information acquired from a subset of potato tubers that are fully visible, it becomes possible to effectively ascertain the quality of the overall harvest, provided that the sample size and frequency are adequate to capture the inherent variability in the crop. Notably, it is worth highlighting that a review of the literature yielded no prior investigations that have considered this approach of automating the sampling process to tackle the challenges encountered during the assessment of tubers in field conditions.

1.3 Research Objectives

Several studies have used machine vision systems for quality grading potato tubers. However, most studies have focused on grading tubers in a controlled environment that requires manual singulation, tuber washing, and artificial lighting, with very few addressing size estimation tasks in field conditions, which the industry considers the most important attribute of interest (Al-Mallahi, personal communication, June 12, 2022). Hence, the overall objective of this study is to develop a new grading technology for the size estimation of tuber crops. The specific objectives are as follows:

i) To develop a method for visually sampling full-size potato tubers flowing on conveyors at the post-harvest stage.

ii) To develop a prototype machine vision system that can be integrated into existing systems for size estimation of potato tubers at the post-harvest conveyors.

Chapter 2: Literature Review

2.1 Potato Production

Potato is a staple food grown in over 160 countries. It is one of the four most important crops in the world (together with wheat, maize, and rice) (Zarzecka et al., 2020), with an annual production of 359 million tonnes in 2020 ("FAO Publications Catalogue 2021," 2021; Potato Facts and Figures - International Potato Center, 2021). In Canada, the potato is the fifth most important agricultural crop (after wheat, canola, soybean, and maize), generating around $1.4 billion in farm cash receipts and $2 billion in potato and potato product exports in 2020 (Potato Market Information Review, 2020-2021 - agriculture.canada.ca, 2022). Additionally, potato is Canada's most frequently farmed vegetable crop, accounting for 28% of vegetable and 16% of horticulture receipts. Most of the all-grown potatoes (approximately 65%) are processed as frozen French fries, potato chips, flakes, and other dried products (Potato Production in Canada – Canadian Horticultural Council, 2019). Therefore, specific quality characteristics, such as texture, defects, tuber size and shape, which are crucial in potato processing, must be assessed by growers (Struik et al., 1990; Si et al., 2017).

Potato tubers are available in various sizes to suit the consumers' demands. Small potato tubers are favoured for planting because they yield more stems per kilogramme; however, large potato tubers are more profitable for producing chips and fries since they yield more kilogrammes per tonne (Abong et al., 2009; Marwaha et al., 2010; Potato World magazine, 2015). Furthermore, Farhadi & Ghanbarian (2014) explained that assessing potato tubers by mass was unconventional due to the mass assessment's relatively slow pace and high cost. Many research studies have established a substitute technique for mass-based potato yield estimation, such as Tabatabaeefa's (2002) finding of a substantial

correlation between minor diameter and mass and Su et al.'s (2018) experiment that showed high accuracy for potato mass prediction based on volume. Thus, estimating potato size, precisely the minor diameter and volume, is critical for growers to maximize profit, assess crop yield, and prepare for post-harvest logistics and marketing.

Sizing potato tubers is a common practice done by visual assessment or passing the harvested crop through V-shaped sieves of different sizes (Mechanical Potato Graders, 2020; Dattatraya et al., 2013). When visual inspection is used, inconsistencies occur due to varying perceptions and fatigue of the human eye. Besides its inconsistency, manual sizing of potato tubers is time-consuming, labour-intensive, expensive, and may be easily impacted by the immediate environment (Narvankar et al., 2005; Razmjooy et al., 2012). As a result, several studies have been conducted to automate the process of potato size grading to boost production speed, accuracy, and efficiency while lowering production costs (Elmasry et al., 2012). One of the earliest trials was proposed by Verma & Kalkat (1975), who designed a potato sizer with an increasing-pitch rubber spool. The prototype potato sizer was a conveyor with a rubber spool and two driving rollers with helical grooves. The performance of the expanding rubber spool potato sizer was tested at varying bed speeds. By slowing the feed elevator conveyor to 45 rpm and using a larger pulley, clustering and occlusion were removed at the start of the size bed. Most recently, Huda et al. (2019) engineered and fabricated a potato mechanical size grader that comprised a hopper, grading unit, prime mover, and catchment tray. The grader was used to classify potato tubers into big (>55 mm), medium (40 – 55 mm), and small (<40 mm) sizes based on three holes in the grading unit. The proposed system had an efficiency of 91.57% and a capacity of 420.10 kg per hour. While mechanical graders are more efficient than visual

inspection at sizing, damage to the potato tubers caused by abrasion of the surfaces is significant, especially at higher speeds (Valentin et al., 2016). Moreover, mechanical graders are expensive and require extensive modification of existing systems.

Machine vision eliminates visual discrepancies and increases sorting efficiency by providing consistent judgement based on estimated parameters (Quilloy & Bato, 2015). Additionally, it offers a non-destructive method for assessing crops. Thus, the past few decades have seen researchers focus on its use for agricultural operations (Tian et al., 2020).

2.2 Machine Vision Systems in Agriculture

Machine vision systems use cameras and computers to replace human visual sense and judgement. Machine vision-based systems generally have three stages: image capture, image processing, and input/output control (Ji et al., 2009). Image processing involves establishing studies and algorithms to evaluate and extract information from a video or still image about an observed object or set of items. Many imaging methods have been used to develop machine vision systems, each with benefits and drawbacks. Black-and-white imaging, colour imaging, stereovision, and hyperspectral imaging are notable examples used in previous research. However, the performance of machine vision systems relies not only on the camera type but also on the algorithm, source of light, and object of interest.

2.2.1 Imaging Systems

Earlier studies have used black-and-white cameras to identify fruit based on reflectance, geometric, and surface aspects (Plá et al., 1993). A black-and-white camera was used to identify scars, fissures, and spreading tips on asparagus plants (Rigney et al., 1996). These basic systems recognized fruits with a high accuracy utilizing a mix of shape, texture, and

light absorbance features. However, Sites & Delwiche (1988) showed that with the addition of colour filters, the contrast between the object and the background was enhanced, leading to higher accuracy of fruit detection.

Numerous aspects of agricultural crop production use the colour camera as a sensing element since it adds layers of information to black-and-white imaging and is relatively cheaper than other imaging systems. Colour imaging has been proven to be effective for quality grading (L. Deng et al., 2017; Firouzjaei et al., 2018), fruit identification (Prasetyo et al., 2020), ripeness detection (Wan et al., 2018), yield prediction (Aggelopoulou et al., 2011; Q. Wang et al., 2013), and weed sensing (Tang et al., 2000). Throop et al. (1993) used colour spectral information to estimate the bruise level in apple fruits. According to the study's results, colour distinctions successfully distinguished injured from healthy tissue. Kataoka et al. (2003) also used a colour camera to estimate the growth status of crops from covered vegetation. Colour images have also been used in developing algorithms for counting marigold flowers (Sethy et al., 2019), kiwifruit (Fu et al., 2019), and apples (Chen et al., 2017). The colour information in an image is crucial for segmentation, although it varies based on illumination. Researchers often do image normalization and convert images from RGB (Red-Green-Blue) to HSV (Hue-Saturation-Value) colour space to reduce the impact of changing lighting conditions on colour images (Garcia-Lamont et al., 2018).

Thermal imaging is a method of converting an object's invisible radiation pattern into visual pictures. Assessment of quality, detection of contaminants, detection of grades, detection of infectious agents, detection of damage, maturity evaluation, and inspection are examples of where thermal imaging is utilized (Sivaranjani et al., 2021). Stoll & Jones

(2007) conducted a feasibility study to determine the practicality of thermal imaging for monitoring stress in grapevines. According to the study, this method could deliver precise and sensitive indications of leaf temperature, which may then be used to calculate stomatal conductance. The use of thermal imaging in agricultural operations throughout the pre-harvest and post-harvest periods as a non-contact, non-destructive technology offers certain benefits, such as working in low-light scenarios. However, thermal imaging has several limitations when compared to other imaging techniques. Costly high-resolution thermal imaging is required, and the accuracy of thermal readings is highly reliant on ambient and meteorological conditions (Ishimwe et al., 2014).

Hyperspectral imaging is widely investigated in agricultural research because the images have an enormous amount of spectral information in addition to the spatial information other types of imaging have (Yud-RenChen & Kim, 2002). Using hyperspectral imaging, Zhao et al. (2008) experimented with detecting whether apples have bruises based on spectral data between 500 nm and 900 nm, which were assessed using principal component analysis. Selected images were then used to eliminate asymmetry in brightness. The research found that their approach correctly identified 88.57% of bruising. This imaging system gives far more helpful information than standard imaging approaches since each image surface pixel has the object's spectral information. Hyperspectral cameras are more costly than other cameras (Mavridou et al., 2019) and challenging to deploy in unstructured terrain without modifications to the existing system.

In addition to conventional vision systems, stereo vision systems provide an alternate method of recognizing objects in three-dimensional (3D) space. These systems use two or more cameras placed a short distance apart to simulate similar binocular vision as in

humans. A notable application of stereoscopic cameras is in the 3D capture of plant structures used in crop and plant monitoring and species-discriminating applications (Lili et al., 2017; Mirbod et al., 2020; Wenhua et al., 2009). Although stereo vision provides the additional benefit of visual inspection in three dimensions, its complexity makes it challenging to use in an unstructured environment, such as an agricultural field.

2.2.2 Computer Vision

Computer vision relates to interpreting, acquiring, and evaluating objects within digital images. It is currently employed in various technologies, including autonomous cars and robots, medical diagnostics, industrial production and surveillance, and remote sensing. Computer vision-based systems are gaining attraction in the food and agricultural industries, notably for quality control. According to Bhargava & Bansal (2018), computer vision enables various farming operations, including land identification, recognition of pest-infested zones, automated categorization, and plant disease detection based on shape, texture, and colour.

Yimyam (2005) developed a system for recognizing, segmenting, and analyzing mango's physical properties. The images were captured using a digital camera and then processed and segmented. The noise in the digital image was eliminated using morphological filtering, and a colour model of the mango samples was developed. This study used structural models to determine the mango's area and shape. The object's colour was also analyzed and classified. Their solution provided a practical alternative to manual sorting. Similarly, to accurately recognize grape berries and detect grape bunches, Andrés et al. (2017) used a visible-light camera to conduct their research. They developed a machine vision system based on geometry and texture using aggregated pixel patches as input. Despite various lighting and partial occlusion circumstances, the system performed well.

Computer vision has also been used to monitor plant growth and detect agricultural disease and nutrient shortage problems. Sannakki et al. (2011) quantified leaf disease using a computer vision-based method. They used a colour camera to capture images of pomegranate leaves and then used image scaling and Gaussian filtering to minimize computational load and eliminate noise. After that, they used K-means clustering to colour-segment the diseased leaf area. Then the diseased area was quantified and graded using basic object-size to pixel-area calibration and fuzzy logic, respectively. In a similar study, Sannakki et al. (2013) diagnosed and classified grape leaf diseases using neural networks. The input data for the proposed system was an image of a grape leaf with a complex background. Green pixels were masked using thresholding, and anisotropic diffusion was employed to remove noise from the image. The disease region was then segmented using K-means clustering. It was shown that the best results were achieved when a feedforward backpropagation neural network was trained for classification.

The popularity of agricultural image processing and computer vision applications has risen as equipment cost has decreased, computing power has improved, and interest in non-destructive food inspection processes has increased (Mahajan et al., 2015). On the other hand, their approaches are limited, and achieving flexibility and stability in a variety of complex situations is difficult. Numerous studies have shown varying degrees of limitation for computer vision applications in the agricultural setting due to the field's unstructured nature.

2.3 Machine Vision for Quality Grading of Potato

Automated quality control of food and agricultural products is quicker and more precise than hand grading and, as a result, has garnered considerable attention. Near-infrared

technology, nuclear magnetic resonance (Chen et al., 1989; McCarthy & McCarthy, 1994), and X-rays were among the first tools for rating the quality of food and agricultural goods (Shahin & Tollner, 1997). While these approaches produced impressive results, their high cost and complexity necessitated the development of a new generation of non-destructive technologies for assessing the quality of agricultural and food goods. Shape classification, defect identification, quality grading, and variety categorization are gaining popularity as applications for automated machine vision systems (Patel et al., 2012).

Noordam et al. (2000) presented a high-speed machine vision system for the quality inspection and grading of potatoes based on size, shape, and external defects - greening, mechanical damages, Rhizoctonia, silver scab, common scab, cracks, and growth cracks. The system utilized Linear Discriminate Analysis (LDA) and multi-layer feed-forward neural network (MLF-NN) techniques for pixel classification. The accuracy of the LDA and MLF-NN sorting techniques for different potato varieties ranged from 86.8% to 98.6% and 88.1% to 99.2%, respectively. In another study, Hassankhani & Navid (2012) developed a machine vision system for sorting potatoes based on size and colour. Their proposed system included a lighting chamber, lighting source, CCD camera, and a computer running hand-crafted feature-based computer vision software. The system achieved an average accuracy of 96.54% across different potato grades. Recent studies include Su et al. (2018), who used 3D imaging for mass estimation, and Shen et al. (2022), who developed an algorithm based on invariant moments, geometrics characteristics, and fractal dimensions for potato shape and size estimation. While these studies have tackled a variety of potato quality grading tasks, they have all been conducted under controlled

conditions with detection simplifying settings such as artificial lighting and manual singulation.

2.4 Application of Deep Learning in Quality Grading of Potato

2.4.1 Overview of Deep Learning

Contemporary scientific studies have witnessed the rising popularity of deep learning and neural networks due to their ability to learn complex scenarios from context (Alazab et al., 2020; Gadekallu et al., 2020). Deep neural networks (DNNs) enable computational models with multiple processing layers to learn data representations with varying degrees of abstraction. As data flows through the network, lower layers learn simple features close to the input data, while higher layers learn more sophisticated features derived from lower-layer features (Shinde & Shah, 2018). The architecture creates a hierarchical and powerful representation of features, making deep learning well-suited for evaluating and extracting meaningful knowledge from massive amounts of data and data obtained from various sources (Zhang et al., 2018). These technologies have significantly advanced the state-of-the-art in speech recognition, visual object recognition, object detection, and various other fields such as drug development and genomics (LeCun et al., 2015). However, deep learning algorithms require extensive data and massive computational power (Bhattacharya et al., 2021).

The hierarchical structure of deep-learning models and their immense learning power enables them to perform exceptionally well at making predictions and classifications while being flexible and adaptable to various highly sophisticated data processing tasks (Pan & Yang, 2010). With the robust capability of automatic feature learning, deep learning methods offer a promising solution for solving complex problems in agriculture, such as

variety recognition, yield estimation, quality grading, growth monitoring, and disease detection (Yang & Xu, 2021).

Among the deep learning methods, convolutional neural networks (CNNs) are a method that can learn complex image features, enabling image classification and recognition in complex scenarios. Therefore, CNN is preferred for computer vision tasks such as image and video recognition. Figure 2.1 illustrates the layers of a CNN as a sequence of convolutional and subsampling layers followed by a fully connected layer and a normalizing (e.g., softmax function) layer. Each layer in a series of several convolution layers captures increasingly complex information as the layers proceed from input to output. The typical computer vision tasks in agriculture are object classification, object detection, and image segmentation, and the architecture of CNN is influenced by the nature of the visual task.

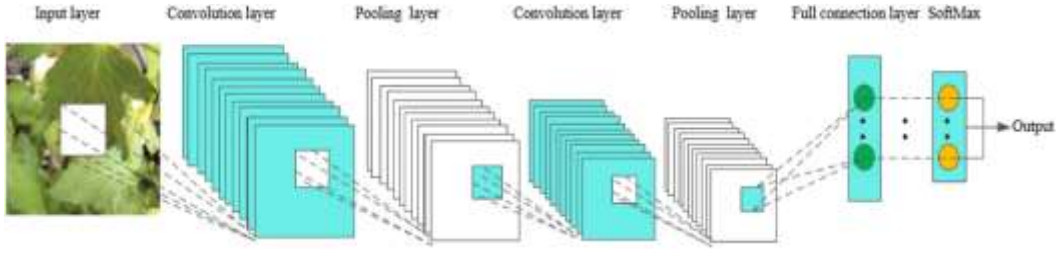


Figure 2.1: Basic structure of CNN (J. Liu & Wang, 2021)

There are several effective and popular CNN architectures upon which researchers might base their models rather than beginning from scratch. Among them are AlexNet (Krizhevsky et al., 2012), VGG (Simonyan & Zisserman, 2014), GoogleNet (Szegedy et al., 2014), and Inception-ResNet (Szegedy et al., 2016), which are used for object classification and recognition. Other popular state-of-the-art architectures are Mask-RCNN (He et al., 2017), FCN (Shelhamer et al., 2014), SegNet (Badrinarayanan et al., 2015), and U-Net (Ronneberger et al., 2015) for image segmentation. For objection

detection, architectures such as YOLO (Redmon et al., 2015), RCNN (Girshick et al., 2013) and SSD (W. Liu et al., 2015) are known to have good performance. Each architecture suits different situations, and choosing the right one is crucial (Canziani et al., 2016). To train CNN models, datasets such as ImageNet (J. Deng et al., 2010) and COCO (T. Y. Lin et al., 2014) are often used for pre-training. Pre-training the models using these datasets allows for transfer learning, which has enabled training models with small to medium-sized datasets and saves training time.

2.4.2 Deep Learning in Potato Quality Grading

Deep learning algorithms have gained attention in several quality grading tasks. To this end, Chicchón & Huerta (2021) developed a machine vision system using a CNN-based approach for rapid volume estimation of potato tubers. They evaluated the effectiveness of SegNet, a deep learning architecture for image segmentation, compared to traditional threshold-based segmentation approaches for calculating potato volume. The deep learning approach accurately detected up to 99% of the potato tubers and predicted volume with up to 90% accuracy. It also had a processing speed of between 10 and 20 tubers per second and was more robust than the traditional method that required calibration depending on environmental parameters.

Dolata et al. (2021) proposed a yield assessment method to estimate potato tubers' physical dimensions on the harvester using Mask R-CNN deep learning algorithm. They introduced a nonlinear softmax regression model for size estimation, calculating the minimum diameter of potato tubers from the ellipse fitted to its perceived contour. They trained the regression model using simulated data and tested it in a near-real-world setting. Results showed that it was possible to estimate the minimum diameter of potato tubers using a

single image obtained with a consumer-grade RGB camera without additional illumination. The proposed model had an average coefficient of determination of 0.8695 for estimating the minimum tuber diameter compared to the actual tuber minimum diameter. A similar study by Lee & Shin (2020) utilized the Mask R-CNN algorithm to assess the size of tubers on the field surface following excavation by a potato digger. The researchers reported a detection accuracy of 90.8% and an average recall of 93.0%, highlighting the effectiveness of the Mask R-CNN algorithm in addressing challenges such as complex backgrounds, clustering, foreign objects, and ambient light. These findings are consistent with those of previous research, emphasizing the potential of the Mask R-CNN algorithm as a valuable tool for potato grading in various settings. Nonetheless, neither investigation confronts the difficulty of identifying tubers in densely crowded scenes in which tubers mutually occlude each other, a prevalent scenario in multiple locations within potato fields.

Deep learning algorithms have also been applied in machine vision systems for the non-intrusive detection of potato surface defects due to the wide variety of tuber shapes and defects (C. Wang & Xiao, 2021). A study by Pandey et al. (2019) developed computer vision software for size grading and surface defect detection of potatoes using CNNs and image processing techniques. The software utilized U-Net architecture to semantically segment images with 50-60 potato tubers from the background and then apply distance transformation and watershed segmentation to obtain the tuber's skin. The method achieved a minimum size distribution accuracy of 97.2% and a defect classification accuracy of 89% on the validation dataset. Marino et al. (2019) used a weakly-supervised learning strategy to classify, localize, and segment potato defects by capturing images of

healthy and defective potatoes in a laboratory environment. They used state-of-the-art CNN architectures to classify defects and a coarse-to-fine segmentation method to determine the precise position of the defect. When evaluated using a multi-label multi-class dataset, the system achieved an average precision of 91% and an average recall of 90%. A recent study by Arshaghi et al. (2021) used CNNs and machine learning to accurately classify potato surface defects, using 5000 images from different sources. Their CNN network included convolutional and fully connected layers, achieving 100% accuracy in classifying surface defects, surpassing typical machine learning techniques. Table 2.1 summarizes some studies that have applied deep learning in the quality grading of potato tubers in the last 5 years.

Table 2.1: Applications of deep learning for quality grading of potato

S/N	Application	Architecture used	Reported performance	Reference
1	Mass estimation	YOLOv5 and DeepSORT	Accuracy: 95.2% RMSE error: 9%	Jang et al. (2023)
2	Disease identification	MobileNet V2 with LSTM and GRU	Accuracy: 99%	Faria et al. (2023)
3	Classification of diseases	CNN and Modified SVM	Accuracy: 99%	Samatha et al. (2023)
4	Identification of hollow heart	ResNet50 and Shallow FCN	AUC: 90% Precision: 90% Recall: 90%	Abbasi et al. (2021)
5	Rapid estimation of tuber volume	SegNet	Accuracy: 90%	Chicchón & Huerta (2021)
6	Tuber size estimation	Mask RCNN Nonlinear regression model	Quality: 88.9% Determination coefficient: 0.869	Dolata et al. (2021)
7	Surface defect classification	SSD Inception V2, ResNet101, and others	Accuracy - 98.7%	Wang & Xiao (2021)
8	Classification of surface defects	CNN	Accuracy - 100%	Arshaghi et al. (2021)
9	Quality grading based on size and appearance	CNN and softmax regression (SR)	Accuracy: 86.6% Loss: 0.304	Su et al. (2020)
10	Size estimation of potato tubers	Mask R-CNN	Accuracy: 90.8% Recall: 93.0%	Lee & Shin (2020)
11	Surface defect classification includes damaged tubers.	AlexNet, VGG-16, GoogLeNet, and others	Precision: 92% Recall: 91%	Marino et al. (2019)
12	Size grading and surface defect detection	U-Net and VGG16	Accuracy: 97.2% for size; 89% for defects	Pandey et al. (2019)

2.5 Summary

Automated quality assessment of food and agricultural products through machine vision is gaining prominence due to its efficiency and precision. Machine vision systems can be employed for diverse applications such as shape classification, defect identification, size grading, and variety categorization. Recent research efforts have explored the use of different imaging systems for various grading concerns. Image features, such as invariant moments, geometric characteristics, edge-based features, and fractal dimension, have been utilized for potato grading tasks in controlled environments and field conditions. Nonetheless, most proposed solutions that leverage handcrafted image features are tailored to address specific challenges that do not generalize to a wide range of variations, such as ambient light, clustering and occlusion, and the presence of foreign objects (Chen et al., 2021).

Deep learning algorithms such as SegNet and Mask R-CNN have gained considerable attention in various quality grading tasks in recent years. These approaches have performed well in outdoor field conditions characterized by spatial and temporal variations. For instance, they have successfully detected surface defects on potato tubers, predicted volume, and estimated the minimum diameter of potato tubers with high accuracy. However, only a few studies have conducted field experiments to assess the feasibility of their proposed solutions in field conditions. Furthermore, the challenge of deep occlusions, where more than half of an object is not visible, which is prevalent at different sites in potato fields, has not been adequately addressed in any of the studies.

Chapter 3: Sampling full-size potato tubers in large-scale clusters

3.1 Introduction

Machine vision has emerged as a prominent tool for evaluating agricultural yield, primarily due to its cost-effectiveness (Ismail & Malik, 2022). However, adopting machine vision is contingent upon minimizing supplementary expenses that may arise from extensive modifications to existing systems. As such, rather than suggesting a mechanical overhaul of the potato post-harvest conveyor system, this research was based on obtaining quality information of potatoes as they pass under a camera by sampling potato tubers on the post-harvest conveyor. This research can be considered successful as long as the sampling size is larger than what can be achieved by manual sampling. This chapter describes the software method for sampling potatoes on the conveyor on-the-go at different clustering scenarios.

3.2 Overview of the sampling method

The method encompasses a series of sequential procedures, commencing with detecting potato tubers in images to ensure the selection of fully visible tubers. For this purpose, instance segmentation of the images is performed using the Mask R-CNN deep learning algorithm. Subsequently, image features are extracted by computing Hu invariant moments and colour and edge-based image features. These extracted features are subsequently utilized to select either one or five tubers per image frame. Figure 3.1 provides an overview of the proposed methodology.

Sampling comprises two primary stages: instance segmentation and tuber sampling, both of which are reliant on supervised machine learning. As such, the availability of annotated datasets is critical for effective model training. In this regard, images of potato tubers captured under laboratory and field conditions were manually labelled to develop the

requisite datasets. The dataset of images obtained under laboratory conditions was employed for training and evaluating the Mask R-CNN model, whereas the datasets aimed at developing and assessing the sampling techniques were created by meticulously labelling each potato tuber depicted in the images as either fully or partially visible. Finally, the preferred sampling technique was evaluated using images obtained under field conditions to gauge its feasibility and applicability in practical scenarios.

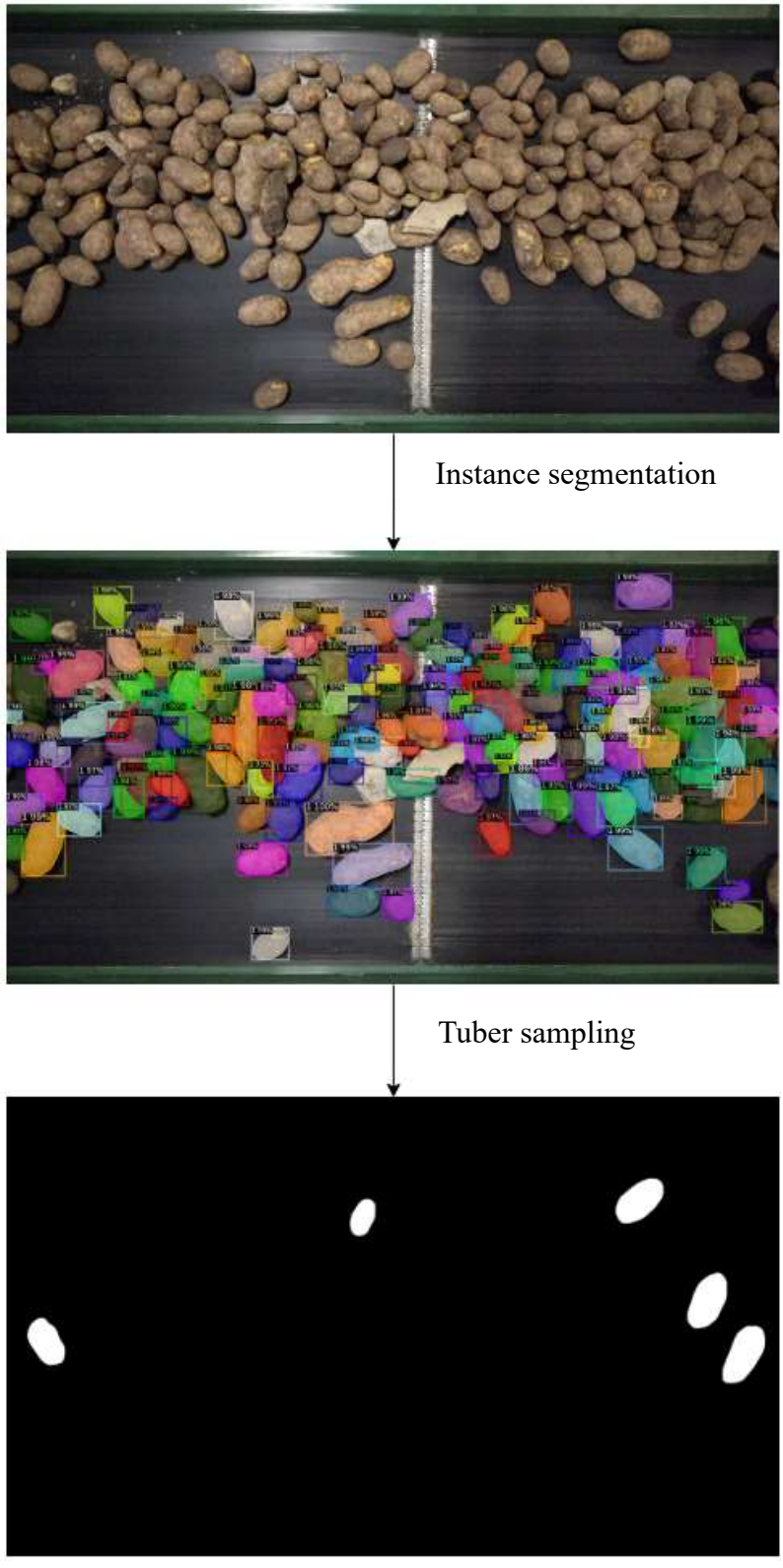


Figure 3.1: Detecting and selecting five tubers from a cluster of potato tubers flowing on a conveyor in the lab. An intermediate process of segmenting all tubers before choosing fully visible based on Mask R-CNN is implemented on any cluster type.

3.3 Image Acquisition and Dataset Creation

The images were taken using an RGB camera (DSC-RX0M2, Sony, Tokyo, Japan), which is cost-effective and easy to use. Additionally, it is a small, rugged, high-resolution RGB camera resistant to vibration, dampness, and shock. These properties made it a good choice, given that it was also used outdoors in the field on the post-harvest conveyor.

High-Definition images (1920×1080 pixels) were captured at the Engineering Laboratory, Faculty of Agriculture, Dalhousie University, and McCain's Farms of the Future, Riverbank, New Brunswick, Canada. The outdoor images were taken at various times of the day during the potato harvesting season in 2021 and 2022, encompassing morning, afternoon, and evening, to ensure that varying lighting conditions in the field are captured. To achieve this, 60 minutes of video data were captured for prevailing circumstances, while 15 minutes were captured for rare cases (see Appendix A-1).

VGG annotator (Dutta & Zisserman, 2019) was used to annotate 36 images obtained in the Laboratory, which comprised three clustering conditions: dense, moderate, and sparse, as shown in Figure 3.2. Every potato tuber was annotated, whether fully or partially visible. In the dense, moderate, and sparse clustering conditions, the average number of tubers per frame was 148, 79, and 54, respectively. Additionally, 54 images obtained outdoors from post-harvest conveyors at the Farms of the Future were annotated to evaluate the proposed methodology under field conditions, incorporating the various clustering scenarios and three lighting conditions: sunny, cloudy, and shaded. For the outdoor images, the average number of tubers per frame was 231, 116, and 53 for dense, moderate, and sparse clustering scenarios, respectively, across all lighting conditions. To ensure its applicability across the two main potato cultivars for French fries and chips, the software was

developed using the Caribou potato cultivar under controlled indoor conditions and evaluated using the King Russet cultivar in field settings. Table 3.1 shows the dataset, which includes information regarding the number of images for different clustering and lighting scenarios and data partitions.

(a) (b) (c)

Figure 3.2: Images of the tubers acquired on laboratory conveyor belts showing (a) sparse clustering, (b) moderate clustering, and (c) dense clustering scenarios.

Table 3.1: Datasets used for the proposed method showing the number of images, clustering and lighting scenarios, and the data split.

Lighting conditions	Variety	Data split	Clustering scenario		
			Sparse	Moderate	Dense
Artificial (Indoor)	Caribou	Training	5	5	6
		Validation	1	2	2
		Testing	5	5	5
Sunny	Russet king	Testing	6	6	6
Shaded		Testing	6	6	6
Cloudy		Testing	6	6	6

3.4 Instance Segmentation

One of the critical problems of this study was identifying an effective technique for segmenting potato tubers from one another (instance segmentation) and the conveyor (semantic segmentation). Image segmentation is necessary before estimating the size of the tubers, as the segmentation quality influences the size estimation accuracy. The Mask R-CNN (He et al., 2017) deep learning algorithm, an enhanced version of Faster R-CNN,

was employed for instance segmentation. The architecture consists of two main stages: the backbone stage and the head stage. The former is constructed using FPN (feature pyramid network) (T.-Y. Lin et al., 2017) and ResNet (He et al., 2016), as well as a Regional Proposal Network (RPN) (Ren et al., 2015) and an ROI (Region of Interest) align layer (Girshick, 2015). This stage proposes regions in the image that might contain potato tubers. The latter consists of fully connected layers where the classification, bounding box and mask predictions take place using the proposed regions of each potato tuber as input from the first stage.

To ensure the reproducibility of the instance segmentation model, the original structure of the implementation was followed utilizing the detectron2 library (Wu et al., 2019), which offers advanced segmentation and detection algorithms. It is an improvement to both detectron and maskrcnn-benchmark whose library is built on PyTorch (Paszke et al., 2019), a widely used deep learning framework based on the Python programming language. This choice of library and framework provides the advantage of seamless compatibility with the remaining components of the computer vision software.

The Mask R-CNN models were trained on a cloud platform offered by the Digital Research Alliance of Canada. The training dataset split of indoor images was used for training the models leveraging available Graphics Processing Units (GPU). To expedite training time and reduce the reliance on extensive training data, three pre-trained weights from the Detectron2 model repository were employed. These weights corresponded to three distinct pre-trained Mask R-CNN models: mask_rcnn_R_101_FPN_3x (based on ResNet-101), mask_rcnn_R_50_FPN_3x (based on ResNet-50), and mask_rcnn_50_C4_3x (based on ResNet-50 backbone with convolution head). The models were adapted to the task by

leveraging the transfer learning technique. Consequently, this research compared the three models in terms of their effectiveness in detecting and segmenting potato tubers.

3.5 Image Feature Extraction

The extraction of valuable features from the detected tuber instance is critical for differentiating fully visible tubers from partially visible ones. In this study, 14 image features were extracted from each tuber instance detected by the Mask R-CNN model to find the appropriate set of features for tuber selection. 7 out of the 14 features were the Hu moments of the image, while the others were based on edge and colour image processing.

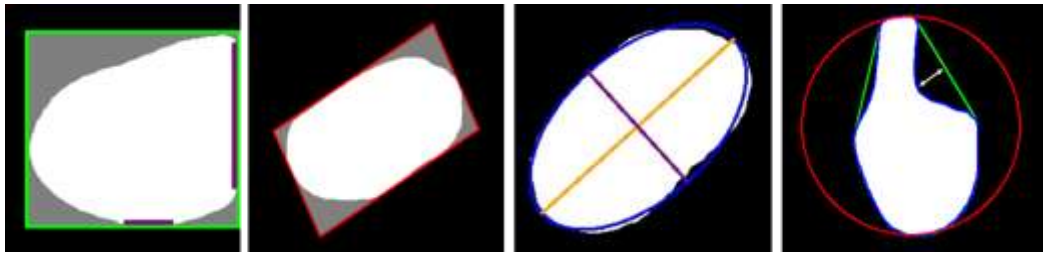
3.5.1 Hu Moments

Image moments are used in computer vision to characterize the shape of an object within an image. The properties of a target can be represented through moments of different orders. Basic shape characteristics can be captured by low-order moments, while high-order moments can provide detailed and complex information. A more complete and accurate representation of the target's properties can be achieved by utilizing low and high-order moments. Thus, all seven image invariant moments obtained from the HuMoments method (see Appendix C) of the OpenCV Python library were considered.

3.5.2 Colour and Edge-based Image Features

Figure 3.3 illustrates the components employed for the development of image features extracted from the detected potato tuber based on colour and edge-based image processing. The Mask R-CNN model was utilized to derive a binary mask of the detected tuber. In conjunction with the tuber mask, the Mask R-CNN model generates a bounding box (bbox) that encompasses the detected region of the tuber within the image. While the bbox generated by the Mask R-CNN model provides valuable information on the location of the detected potato tuber in the image, it does not offer insights into the tuber's orientation.

An angle-adjusted bbox was introduced to capture the angular projection of the tuber. This adjustment involved precisely fitting a bbox onto the tuber mask, enabling the incorporation of the tuber's orientation into the image analysis. Furthermore, to gain insights into the shape characteristics of the detected potato tuber, an ellipse was fitted to the external contour of the tuber mask. The deviation of the detected potato tuber from an idealized elliptical shape was assessed by fitting an ellipse to the contour. An overview of these handcrafted image features is provided below:



(a) (b) (c) (d)

Figure 3.3: Detected tuber masks (white) showing (a) avg_val_bbox, which are the purple straight lines on the detected bbox (red) and S/N_1, which is the ratio of grey vs white pixels; (b) S/N_2 which is the ratio of the grey vs white pixels; (c) Ellipticalness and ellipse ratio; (d) convexity defect which is the deviation of the tuber contour (blue) from the convex hull (green) and the circularity which the ratio of the area enclosed by the tuber contour (blue) to the area enclosed by the circle (red).

1. Average pixel count on detected bbox edges (avg_val_bbox): The number of pixels on each of the four edges of the detected bounding box for the tuber mask was computed, and their average was taken. This provided an estimate of the average density of tuber mask pixels along the edges of the bbox. These pixels are highlighted as purple in Figure 3.3 (a).
2. Signal-to-Noise ratio 1 (S/N_1): This ratio is the number of tuber mask pixels divided by the total number of pixels enclosed by the detected bbox (red), including the white and gray pixels shown in Figure 3.3 (a). The signal component

is the number of tuber mask (white) pixels. The outer contour of the mask was extracted by performing binary thresholding with a pixel value of 200 to remove the detected bbox and then applying the findContours method of the OpenCV library to get the outline of the tuber mask, which was then utilized to calculate the number of tuber mask pixels.

3. Signal-to-Noise ratio 2 (S/N_2): This ratio is the number of tuber mask pixels divided by the total number of pixels enclosed by the angle-adjusted bbox (red), as shown in Figure 3.3 (b). The signal is determined using the same approach as that for S/N_1. In contrast, the noise component is obtained by calculating the product of the width and height of the angle-adjusted bbox fitted to the external contour of the segmented tuber mask. The angle-adjusted bbox considers the projection angle of the segmented tuber mask and is generally smaller than the detected bbox; thus, it typically has a higher value than S/N_1.
4. "Ellipticalness": This is the ratio of the area of an ellipse fitted to the external contour of tuber mask pixels to the number of tuber mask pixels. The area of the fitted ellipse is calculated using the ellipse area formula that considers the major (green) and minor (blue) diameters of the fitted ellipse shown in Figure 3.3 (c). The area of the ellipse is calculated using Equation 3.1.

$$A = \pi ab, \tag{3.1}$$

where a and b are the lengths of the major and minor axes of the ellipse fitted to the tuber mask, respectively, and A is the area of the fitted ellipse.

5. "Ellipse ratio": It is the ratio of the number of tuber mask pixels enclosed by the contour of the fitted ellipse to the area of the tuber mask. It differs from

"ellipticalness" since it uses the number of tuber mask pixels enclosed by the fitted ellipse instead of the area of the ellipse.

6. Circularity: This parameter evaluates how closely the shape of a segmented tuber resembles that of a circle. The circularity (*C)* was calculated by using the estimated area (*S*) and perimeter (*P*) of the segmented tuber mask, using Equation 3.2

$$C = \frac{4\pi \times S}{P^2} \qquad (3.2)$$

7. Convexity defect: This parameter is the deviation of the tuber mask contour (blue) from its convex hull (green), as shown in Figure 3.3 (d).

The handcrafted image features can be divided into two categories: area-based features, namely S/N_1 and S/N_2, and edge-based features, which comprise the five other features that pertain to the shape of the segmented tuber mask. The circularity and convexity defects were derived using the pre-existing functionalities of the OpenCV python library, while custom code was developed using the OpenCV and sci-kit-image python libraries to obtain the remaining features.

3.6 Tuber Sampling Techniques

Using the validation dataset of indoor images, 536 detected tuber masks were obtained, comprising 140 fully visible tubers and 396 partially visible ones. This dataset was utilized for developing two sampling techniques, as discussed in the following subsections.

3.6.1 Threshold-based Sampling

The primary objective was to find the threshold value for each handcrafted feature that distinguishes between fully and partially visible potato tubers. The threshold values were set based on the 25th, 50th, and 75th percentiles of each feature as well as the mean, maximum, and minimum values across the fully and partially visible tubers. Additionally,

the point-biserial correlation was calculated between each independent and the dependent variable to select only those features that could discriminate fully visible tubers from partially visible ones. However, the avg_val_bbox was not employed as a thresholding parameter; instead, it was utilized to sort and select the top one or five tubers that satisfied the thresholding criteria for all features. This was because the fully visible tuber masks typically had fewer pixels on the edges of the bounding box. Moreover, the avg_val_bbox served provided a way to eliminate tubers at the edge of the image that the camera did not see completely. Therefore, if two tubers had similar values for the other features, the one with fewer average pixels on the bounding box edges was preferred for sampling.

3.6.2 Machine Learning-based Sampling

The application of machine learning algorithms can facilitate the identification of the relationship between independent and dependent variables; it could be used to predict whether a given potato tuber is fully visible based on the handcrafted image features. In this study, three machine learning models, namely random forest, support vector machines, and logistic regression, were trained using a 4-fold cross-validation approach for binary classification purposes (i.e., distinguishing between fully and partially visible potato tubers).

Breiman (2001) proposed the random forest (RF) algorithm, an ensemble-based machine-learning technique that utilizes a collection of decision trees. In the RF algorithm, decision trees are trained on randomly sampled subsets of the data, and the resulting predictions are averaged to obtain a posterior class probability. The use of multiple decision trees trained on random subsets of the dataset introduces multiple sources of bias, which effectively counteract the over-fitting challenge commonly encountered in the decision

tree algorithm. The number of decision trees was set to 100 to train the RF classifier, and the maximum depth, which is the longest path between the root node and the leaf node in a decision tree, was set to None (nodes are expanded until all leaves are pure). It is worth noting that selecting a high maximum depth or a low number of decision trees could result in over-fitting, as the model may become overly complex and fail to generalize to new data or suffer from high variance due to a small number of trees.

Logistic regression (LR) is a parametric algorithm that belongs to the family of Generalized Linear Models (GLMs). The algorithm employs a sigmoid function to model the likelihood of the binary target variable. The formula for logistic regression (LR) is given in Equation 3.3. Before training the LR model, a standard scaling technique was applied to normalize the dataset. This technique transforms each feature in the data into a distribution with a mean value of 0 and a standard deviation of 1 using Equation 3.4. Standardizing the dataset is necessary for LR since the L2 (ridge) and L1 (lasso) regularisers assume that all features are centred around 0 and have a comparable variance. Failure to standardize the dataset may result in issues where certain features dominate the objective function due to their higher variance, leading to an estimator that fails to learn correctly from other features. The L1 regularisation (Lasso regression) was used as the penalty as it adds the absolute value of the magnitude of the coefficient to the loss function. This approach helps reduce the model's complexity and avoid over-fitting, thereby ensuring better generalizability of the LR model. As a result, the preference for L1 regularization over L2 regularization (Ridge regression) was based on the prioritization of model simplicity and generalizability.

$$P(Y = 1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p)}} \qquad (3.3)$$

where Y is the binary response variable, X is the vector of predictor variables, β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients for the predictor features.

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{3.4}$$

where z_{ij} represents the normalized value of the feature x_{ij}, which, in turn, denotes the original value of the feature j in the *ith* sample, μ_j is the mean of feature j across all samples, and σ_j is the standard deviation of feature j across all samples.

Support Vector Machines (SVMs) have gained popularity in discrimination tasks due to their ability to effectively combine many features and identify an optimal separating hyperplane based on kernel trick. In the implementation, the default radial basis function (RBF) kernel was used, which is based on Euclidean distance optimization and has the same form as the Gaussian probability density function kernel (Heikamp & Bajorath, 2013), as described in Equation 3.5. Consequently, dataset standardization was done using Equation 3.4 before training the SVM model.

$$\text{RBF}(x, c, \alpha) = \exp\left(-\frac{|x-c|^2}{2\alpha^2}\right) \tag{3.5}$$

where x represents the input vector, c is the centroid of the RBF function, $|\cdot|$ denotes the Euclidean distance, and α is a hyperparameter that controls the "spread" of the kernel and is set as shown in Equation 3.6.

$$\alpha = \frac{1}{n * \sigma^2} \tag{3.6}$$

where n is the number of features in the dataset, and σ^2 is the variance of the data.

The scikit-learn library, a library in Python for predictive data analysis, was leveraged in implementing all three machine learning algorithms. The prediction probabilities generated by the machine learning models for each segmented potato tuber mask were then utilized to select the top one or five tubers per frame.

3.7 Feature Selection

Feature selection was applied to achieve the following objectives: (1) to identify a subset of features capable of effectively discriminating fully visible tubers from partially visible ones, (2) to remove redundant features, and (3) to reduce the processing time required for selecting a subset of fully visible tubers. In this regard, Sequential Forward Feature Selection (SFFS) was employed to increase the handcrafted features' relevance by eliminating redundancy. This procedure entails adding features (forward selection) in a greedy manner to create a feature subset. Based on the cross-validation score, the estimator selects the best feature to add or remove at each stage. This research employed a supervised learning approach based on the RF classifier as an estimator, with the area under the receiving operating curve (AUROC) as the scoring metric (see Appendix D). The SFSS was implemented using the SequentialFeatureSelector method of the machine learning extensions (Mlxtend) Python library.

3.8 Performance Evaluation Metrics

The computer vision metrics of average precision (AP) and average mean intersection over union (mIoU) were employed to evaluate the performance of the three trained Mask R-CNN models based on the validation dataset of indoor images. Subsequently, the best-performing model was assessed across the three clustering scenarios using the test dataset of indoor images.

Furthermore, the test dataset of indoor images was utilized to evaluate the effectiveness of the two proposed tuber sampling techniques. The evaluation was based on two metrics: the average mIoU and the sampling accuracy of the selected tubers in various clustering scenarios. These metrics were used to assess the accuracy and reliability of the sampling techniques in accurately selecting well-segmented, fully visible tubers.

To compare the performance of the three machine learning models trained to discriminate between fully visible and partially visible tubers, the recall, precision, and AUROC were evaluated using a 4-fold cross-validation. The precision used in assessing the machine learning-based sampling technique is unrelated to the AP used to evaluate the Mask R-CNN model, although both are measures of recognition quality; AP for tubers whether fully or partially visible and precision for fully visible tubers.

The AP was employed to assess the object detection proficiency of the Mask R-CNN model, while the average mIoU was utilized to evaluate the segmentation quality. These metrics constitute fundamental evaluation criteria for the task of tuber size estimation. When assessing characteristics such as size, models exhibiting a higher average mIoU are generally preferred as the precision of the parameters becomes increasingly accurate with better segmentation of the potato tubers.

3.9 Results of Instance Segmentation

The Mask R-CNN models were trained for 500 epochs, utilizing the GPU (Tesla 4, 32 GB) available on the Digital Research Alliance of Canada. The training process took approximately 6 hours to complete. Subsequently, the validation dataset consisting of indoor images was used to assess the AP and average mIoU metrics, as presented in Table 3.2. Among the three trained models, the Mask R-CNN model trained on the

mask_rcnn_R_101_FPN_3x pre-trained weight exhibited slightly superior performance compared to the other two evaluated models. This finding aligns with the results of (He et al., 2017) regarding the enhanced performance of the Mask R-CNN model when ResNet-101 is used as the backbone. As a result, the model built on mask_rcnn_R_101_FPN_3x was selected as the most preferred model and carried out further evaluations, focusing on the various clustering scenarios.

Table 3.2: AP, AR, and Average mIoU for all potato tubers for the three Mask R-CNN models using the validation dataset of indoor images.

Weight	AP (%)	AR (%)	Avg. mIoU (%)
mask_rcnn_R_101_FPN_3x	73.16	81.00	90.40
mask_rcnn_R_50_FPN_3x	73.09	82.11	89.50
mask_rcnn_50_C4_3x	71.19	82.06	89.30

Figure 3.4 shows the average mIoU values for various clustering scenarios, indicating a consistent and high segmentation quality achieved by the Mask R-CNN model. However, examining the AP metric, which assesses the model's detection accuracy, reveals a 27% decline in dense clustering compared to sparse clustering. This reduction in detection accuracy can be attributed to the Mask R-CNN model's inability to detect all instances of tubers amidst significant clustering and occlusion challenges. These findings align with the observations made by Dolata et al. (2021), who also reported a decline in the detection accuracy of the Mask R-CNN model when confronted with the upper quarter of the image characterized by significant clustering and occlusion of tubers.

Nevertheless, despite the impact of clustering conditions on the number of tubers detected by the model, it can accurately segment the tubers it detects, as indicated by the relatively high average mIoU values. This consistency in achieving high mIoU is noteworthy since certain quality grading tasks, such as size estimation and shape recognition, require

precise segmentation of the target object. However, as depicted in Figure 3.4, not all detected tubers are fully visible and accurately segmented, presenting challenges in relying solely on the Mask R-CNN model's detections for tuber grading purposes.

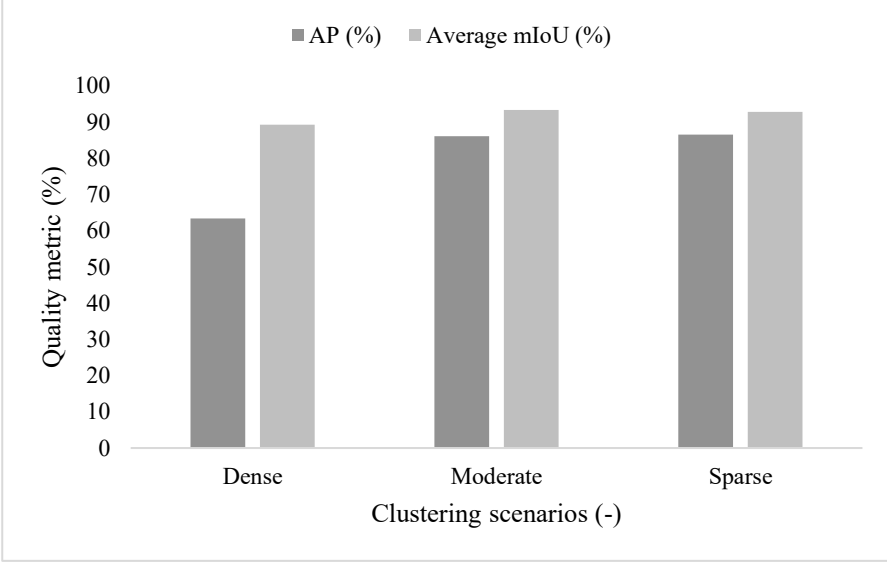


Figure 3.4: AP and Average mIoU of Mask R-CNN model based on different clustering scenarios using the test dataset of images acquired in the Laboratory.

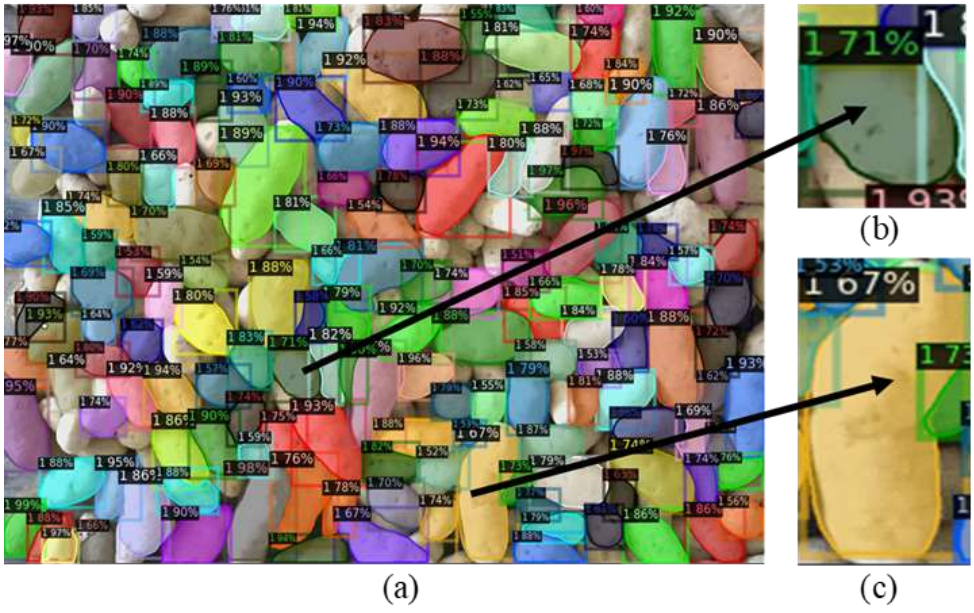


Figure 3.5: (a) Instance segmentation of a densely clustered image obtained from the laboratory showing (b) a partially visible tuber that appears elliptical and (c) poorly segmented tubers (two tubers detected as one).

3.10 Results of Sampling

Two sampling techniques were employed to address the challenges associated with detecting tubers in densely clustered scenarios: thresholding and machine learning. They were applied to select a subset of fully visible tubers among those detected by the Mask R-CNN model. The following sections present the results derived from applying these sampling techniques, accompanied by a rationale for selecting one technique over the other.

3.10.1 Threshold-based Sampling

In threshold-based sampling, the threshold values of handcrafted image features were computed by calculating their mean, range, 25th, 50th, and 75th percentile values for both fully visible and partially visible tuber classes. Furthermore, the point biserial correlation coefficient was computed between each standardized feature and the target variable to identify the features for tuber selection. Accordingly, the avg_val_bbox was used as a sorting parameter to select a subset of tubers meeting the defined threshold criteria.

Based on the distribution of each feature and the correlation coefficient, seven features were selected as thresholding parameters; two Hu image moments (H3 and H4) and five colour and edge-based image features (S/N_1, S/N_2, ellipticalness, circularity, convexity defect). These parameters were chosen because they had a correlation coefficient of 0.2 or higher with the dependent variable with acceptable statistical significance (p-value < 0.05). The point biserial correlation coefficient was utilized because the dependent variable was naturally binary, while the independent variables were continuous. Moreover, this correlation coefficient is robust against slight deviations from normality. After the threshold values were determined for the handcrafted features, the test dataset of images acquired in the Laboratory was used to evaluate the performance of the

sampling technique based on the sampling accuracy and the average mIoU of choosing one or five tubers per frame. The results are presented in Table 3.3.

Table 3.3: Sampling accuracy and Average mIoU for the threshold-based sampling of indoor images based on selecting 1 and 5 tubers per image.

	Sampling 1 tuber per image		Sampling 5 tubers per image	
	Tubers Fully visible/ Sampled	Avg. mIoU (%)	Tubers Fully visible/ Sampled	Avg. mIoU (%)
Dense	5/5	91.67	15/19 ± 0.09	92.40
Moderate	5/5	93.89	21/21 ± 0.00	93.33
Sparse	5/5	95.30	12/13 ± 0.13	94.20

The utilization of threshold-based sampling yielded a sampling accuracy of 100% in the selection of a singular tuber per frame. Consequently, this sampling rate translates to an approximate sampling size of 0.7% of the overall harvest, based on the dense scenario, at 1fps. This represents a possible increase of 3400% compared to the current practice observed at McCain's Farm of the Future, where less than 0.02% of the harvested tubers are selected for grading the yield. Furthermore, the average mIoU consistently remained high when one or five tubers were sampled per frame, indicating successful segmentation of the chosen tubers and their potential suitability as representative samples for grading. Nevertheless, when choosing five tubers per frame, the sampling accuracy of the threshold-based approach decreased by 21% (accuracy of 78.95% with a 95% confidence interval of 69.99% to 87.91%) in the densely clustered scenario compared to the moderately clustered one, where 100% sampling accuracy was achieved. Also, when five tubers were selected, the sampling technique selected 48 tubers across 15 frames, which is 64% of the expected sample size of 75 (5 tubers per frame for 15 frames). However, this would not cause problems as the rigorous threshold values ensure the selection of only fully visible tubers when only a single tuber is selected per frame.

3.10.2 Machine Learning-based Sampling

The challenges of limited sample size and reduced sampling accuracy associated with using fixed thresholds resulted in exploiting machine learning algorithms. These machine-learning models were all implemented using scikit-learn, a library in Python for predictive data analysis. Initially, all 14 features were used for training the models. The results of the performance of the trained models are summarised in Table 3.4.

Table 3.4: Comparison of machine learning models for sampling based on the average of 4-fold cross-validation using the validation dataset on all features.

Machine learning model	Recall (%)	Precision (%)	F1-Score (%)	AUROC (%)
Random Forest	82.81	84.11	83.46	95.13
Support Vector Machines	70.83	73.90	72.33	88.50
Logistic Regression	71.35	73.42	72.37	88.80

The random forest classifier performed best in all evaluated classification measures compared to the other two classifiers. The discrimination capability of a model can be assessed using the AUROC. An AUROC of 50% indicates an inability to distinguish fully visible tubers from partially visible ones, while values between 70% and 80% are considered acceptable, 80% to 90% are excellent, and values exceeding 90% are remarkable (Mandrekar, 2010). Considering these criteria, all three models performed well across the various clustering scenarios. b

Nevertheless, utilizing all 14 features for sampling may not be desirable due to potential redundancies and increased processing time, as explained in Section 3.7. Therefore, sequential forward feature selection was employed to identify a subset of features with the highest predictive power using the random forest classifier as the estimator. The individual contribution of each feature to the estimator calculated as the decrease in node impurity weighted by the probability of reaching that node is presented in Figure 3.6. Furthermore, Figure 3.7 illustrates the relationship between the AUROC and the number

of features, demonstrating how the AUROC varies with the inclusion of additional features.

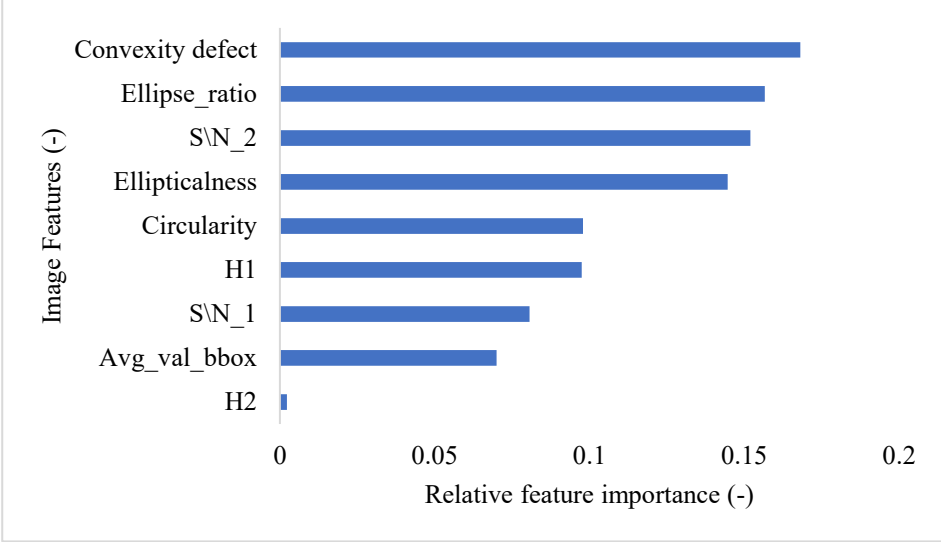


Figure 3.6: Relative feature importance of all handcrafted features obtained from the random forest model.

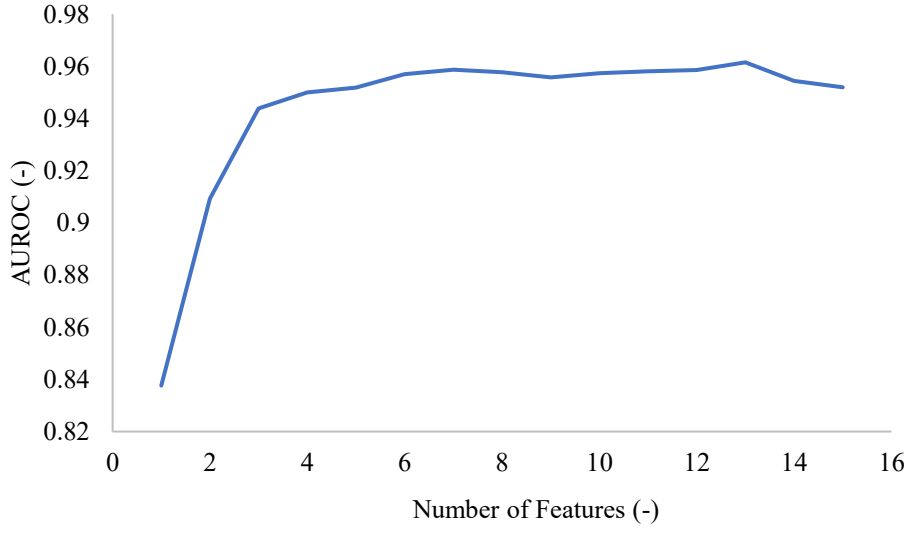


Figure 3.7: AUROC variation with the number of features in the training dataset.

The results presented in Figure 3.6 shows that the convexity defect, which quantifies the deviations of the tuber contour from its convex hull, had the highest contribution to the random forest model, accounting for 16.8% of the relative feature importance. Additionally, the combined influence of features related to the resemblance of the tuber

mask (ellipse_ratio and ellipticalness) to an ellipse proved to be the most predictive in distinguishing between fully visible and partially visible tubers, contributing a total of 30% relative feature importance. Despite the conceptual relationship between the ellipse_ratio and ellipticalness, their Pearson correlation coefficient was estimated to be -0.14, indicating minimal association. However, the Hu moment features exhibited low predictiveness, a similar insight obtained using the point biserial correlation coefficient for feature selection in the threshold-based technique.

Furthermore, the analysis depicted in Figure 3.7 reveals that there is only a marginal increase in the AUROC when the number of features surpasses six. Therefore, to optimize the AUROC, the SequentialFeatureSelector method from the Mlxtend library in Python was employed to select an optimal combination of features, including the convexity defect, ellipse_ratio, ellipticalness, circularity, avg_val_bbox, and S/N_1 as the most predictive features. The results of using these selected features for training the three machine learning models are shown in Table 3.5.

Table 3.5: Comparison of the machine learning models for sampling based on the average of 4-fold cross-validation using the validation dataset on selected features.

Machine learning model	Recall (%)	Precision (%)	F1-Score (%)	AUROC (%)
Random Forest	85.42	84.61	85.01	95.29
Support Vector Machines	73.45	66.67	69.90	86.75
Logistic Regression	68.75	74.89	71.69	88.01

Comparing the performance of the machine learning models using the complete set of 14 features (Table 3.4) with the selected subset of features (Table 3.5), it is observed that the differences in performance are marginal. The random forest classifier performed slightly

better when only seven features were used. This can be attributed to the removal of redundant features, leading to improved generalization and a reduced risk of overfitting. As a result, the model achieved better performance when applied to unseen datasets.

Among the evaluated models, the random forest classifier trained on a subset of features outperformed the others, exhibiting the most favourable performance. Consequently, this classifier was selected as the preferred choice for further development and testing. The sampling accuracy and average mIoU were evaluated to assess its effectiveness based on the test dataset of indoor images. The specific results of this evaluation can be found in Table 3.6.

Table 3.6: Sampling accuracy and Average mIoU for random forest-based sampling using the testing split dataset of indoor images based on selecting 1 and 5 tubers per image.

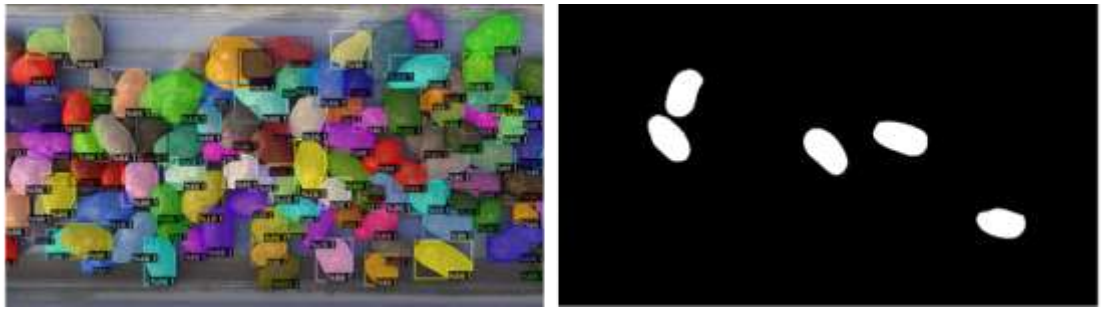
	Sampling 1 tuber per image		Sampling 5 tubers per image	
	Tubers Fully visible/ Sampled	Avg. mIoU (%)	Tubers Fully visible/ Sampled	Avg. mIoU (%)
Dense	5/5	93.37	24/25 ± 0.05	93.11
Moderate	5/5	94.87	24/25 ± 0.05	94.47
Sparse	5/5	96.16	25/25 ± 0.00	95.18

As seen in Table 3.6, implementing the random forest model for sampling purposes resulted in a higher average mIoU across the various clustering scenarios than that obtained using the threshold-based technique. Furthermore, since the sampling technique involved selecting the first few elements in an array sorted in descending order by the model's prediction probabilities, the number of samples obtained was equivalent to 100% of the expected selected tubers, resulting in a 56% increase in the sampling size when compared with the threshold-based technique. Moreover, the model had a 21.61% improvement in sampling accuracy (with a 95% confidence interval of 90.9% to 100%) in the dense scenario compared to threshold-based sampling. For the grading of potato tubers,

the greater the sample size and sampling accuracy, the more accurately the samples represent the population; hence, the random forest model was chosen for in-field evaluation over the threshold-based technique.

3.11 Sampling in Field Conditions

To test the algorithms outdoors, the test dataset of images acquired at the post-harvest conveyor at the farms of the future during the 2022 harvest was used to assess the random forest-based sampling technique in terms of the average mIoU and the sampling accuracy. Figure 3.8 shows the result of the instance segmentation and sampling of densely clustered potato tubers in cloudy scenario, while Table 3.7 summarises the results for all scenarios evaluated using the random forest model for sampling.



(a) (b)

Figure 3.8: Densely clustered image in a cloudy scenario showing (a) tuber segmented image and (b) 5 fully visible tubers sampled from those detected.

Table 3.7: Average mIoU and sampling accuracy for different lighting and clustering scenarios using the dataset of images acquired from the field when 5 tubers are sampled.

	Sunny		Shaded		Cloudy	
	Tubers Fully visible/ Sampled	Avg. mIoU (%)	Tubers Fully visible/ Sampled	Avg. mIoU (%)	Tubers Fully visible/ Sampled	Avg. mIoU (%)
Dense	26/30 ± 0.11	92.23	23/30 ± 0.16	93.53	26/30 ± 0.11	93.77
Moderate	27/30 ± 0.10	91.47	27/30 ± 0.10	92.68	30/30 ± 0.00	93.88
Sparse	29/30 ± 0.05	93.23	27/30 ± 0.10	93.09	30/30 ± 0.00	92.84

The results in Table 3.7 demonstrate that the sampling accuracy exhibits higher variance than the average mIoU. Specifically, the shaded and dense condition yielded the most

unfavourable sampling accuracy, which was 23.33% lower than the optimal case, where the accuracy was 100%. In contrast, the worst average mIoU was only 2.6% lower than the best case, achieving the lowest value of 91.47%. This disparity between the sampling accuracy and average mIoU can be attributed to the Mask R-CNN model's ability to precisely segment fully and partially visible tubers.

Despite the observed variability in sampling accuracy, the lowest value obtained was still relatively high at 76.67% (with a 95% confidence interval of 66.17% to 87.27%), corresponding to the successful sampling of approximately four fully visible tubers out of five sampled per frame. Notably, across all scenarios, the sampling accuracy and mIoU exceeded 90%, demonstrating the applicability of the proposed sampling approach in the field.

The significance of sampling is evident in the dense clustering scenario. However, further investigation reveals that the approach is also beneficial in the sparse scenario, where the Mask R-CNN model occasionally detects the shades (yellow bbox) of the tubers as potato tubers, as illustrated in Figure 3.9. This further emphasizes the importance of the proposed method in ensuring that only actual potato tubers are selected for tuber assessment. Moreover, the method can enhance the size assessment of tubers by ensuring that the sample size and frequency capture the variability in harvested tubers. Additionally, utilizing a machine learning approach for sampling provides flexibility in increasing the sample size by varying the number of tubers selected per frame.

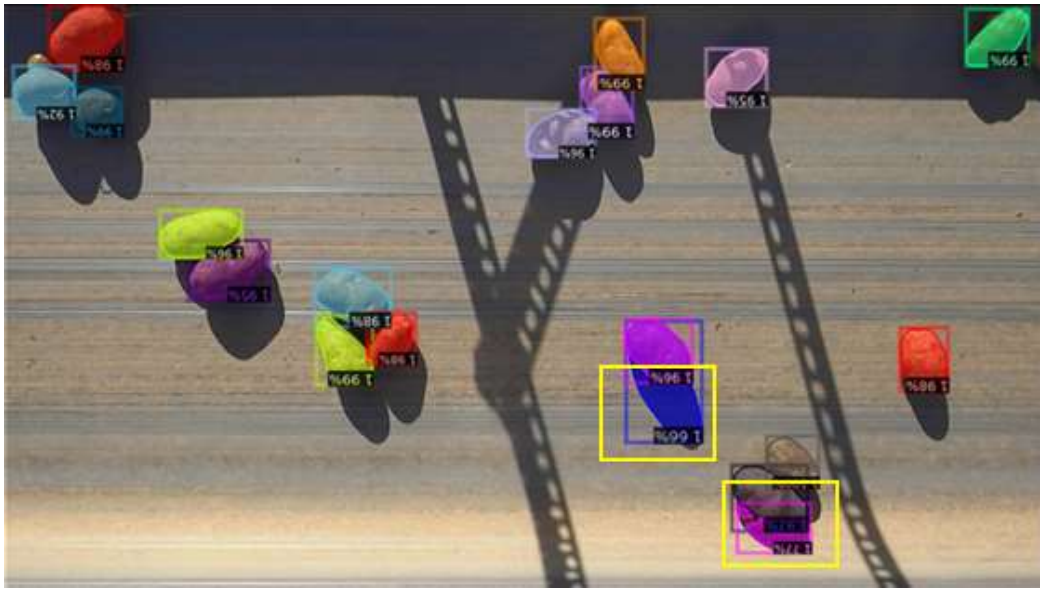


Figure 3.9: Segmented image in sparse with shade scenario showing detection of tuber shade (Inside the yellow boxes) by the Mask R-CNN model.

3.12 Conclusion

The results in this chapter demonstrate the ability to sample fully visible potato tubers -- the first step in the quality grading process. The proposed method utilizes the Mask R-CNN deep learning algorithm and an image feature-based machine learning model to accurately segment images of potato tubers and select up to five fully visible tubers per frame. The method demonstrates a sampling accuracy of 90.74% and an average mIoU of 93% in real-world scenarios, even under challenging conditions such as varying lighting and clustering. These results mean that the physical quality attributes obtained from the sampled tubers can be trusted as they will be taken from fully visible tubers.

Chapter 4: Machine vision system for size estimation

4.1 Introduction

This chapter demonstrates the development of a machine vision system to estimate the lengths of potato tubers using the sampling method described in Chapter 3. The machine vision system was validated in a laboratory setting using static and dynamic conveyors, demonstrating its accuracy and potential to enhance size grading in the potato industry.

4.2 Assembly of the machine vision system

The block diagram in Figure 4.1 depicts the interconnection of all electronic components in the machine vision system. The RGB camera, which served as the sensing element, was connected to a frame grabber (HDR Capture Card, MYPIN, Guangdong, China) that transfers the images to an industrial computer (Nuvo-7006E, Neousys, Taipei, Taiwan). To switch the camera on, a breakout board (Arduino UNO Revision 3, Somerville, USA) as programmed to extend and retract the arm of a linear servomotor (PQ-12-r, Actuonix, British Columbia, Canada) that pressed the power button whenever the computer was turned on. The camera and servo motor were housed in a box (Figure 4.2b), while the computer was housed in another box (Figure 4.2c).

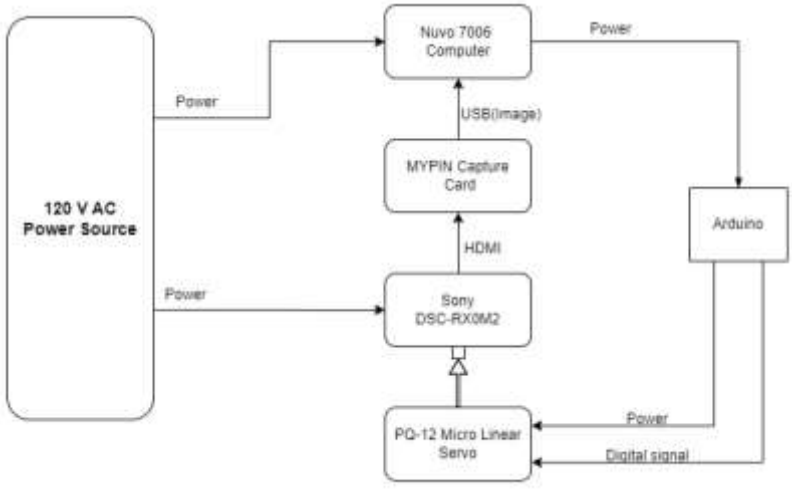


Figure 4.1: Block diagram of the machine vision system

4.3 Indoor apparatus

Figure 4.2 illustrates the configuration of the indoor machine vision system, encompassing the mechanical mount, a series of four conveyors (RCO, New Brunswick, Canada) and the electronic components associated with the imaging system. In Figure 4.2a, conveyors (ii), (iii), and (iv) exhibit uniform characteristics, featuring a length of 4m and a maximum speed of 1.14m/s. These conveyors are driven by 2 hp industrial 3-phase motors (GRA0024D-TC-01, Techtop, British Columbia, Canada). Conversely, the conveyor (i) served as a temporary storage unit for the tubers and has a length of 4.43m. It is powered by a 3 hp industrial 3-phase motor (GRA0024D-TC-01, Techtop, British Columbia, Canada), enabling it to operate at the same maximum speed as the other conveyors. The utilization of all four conveyors enabled the creation of different clustering scenarios by applying varying speeds across each conveyor.

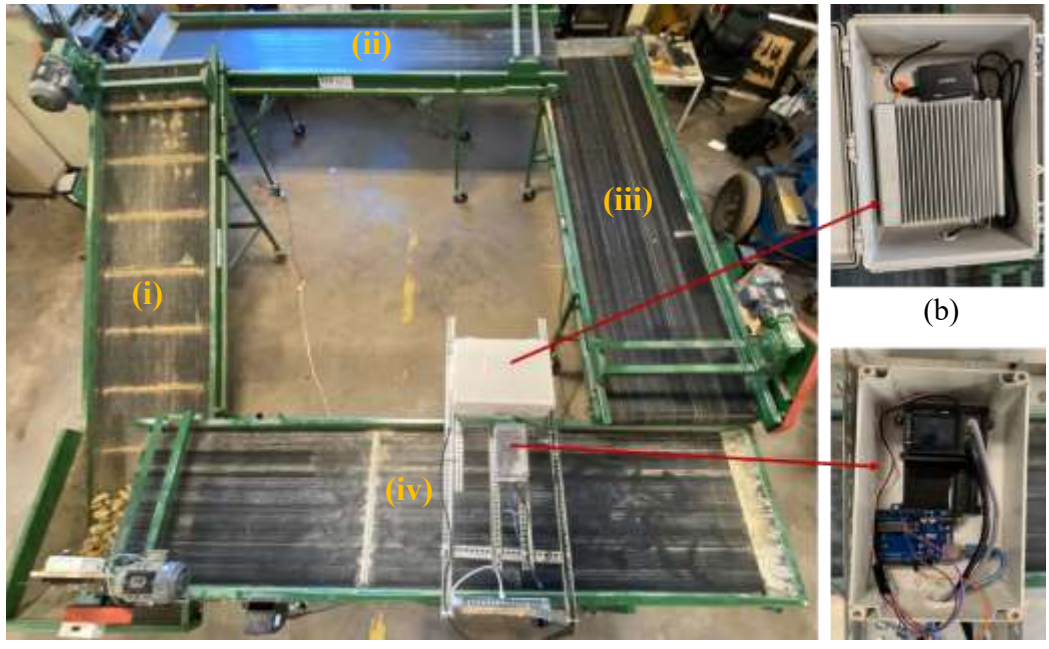


Figure 4.2: Machine vision system set-up in the Laboratory showing (a) the conveying system and the mechanical mount; (b) computer box housing the frame grabber and computer; (c) camera box housing the RGB camera, servo motor, and Arduino UNO.

The electronic components of the machine vision prototype were mounted on a mechanical structure attached to the conveyor (iv) in Figure 4.2a. This mechanical mount is assembled from metallic slotted angles to position the camera 1.1m above the conveyor belt. As a result, the camera's field of view spans a width of 0.8m and a length of 1.3m. Shock-absorbing pads were placed between the conveyor body and the mechanical mount to mitigate the adverse effects of conveyor vibrations.

4.4 Frame Rate Synchronization with Conveyor Speed

The conveyor speed was monitored using the magnetic shaft sensor (590060, Reed, Wisconsin, USA) to adjust the frame rate regularly. The computer read the sensor data serially to enable the adjustment of the camera frame rate based on conveyor speed. This step was necessary to prevent the repetition of frames with the same tubers. The image length per frame was 1.3m, and to avoid tuber repetition in the image, a frame acquisition speed of 1 frame per 2 seconds was used when the conveyor speed was 1.14 m/s, which was the maximum speed. For a conveyor speed of 0.57m/s, a frame rate of 1 frame per 4 seconds was used. At this rate, it is guaranteed that no same tuber will be sized more than once. Equation 4.1 provides the relationship between the conveyor speed (v) and the frame rate (f) used in this study.

$$f = 0.5 \times v \qquad (4.1)$$

4.5 Size Estimation and Calibration

To estimate the major and minor diameters of the fully visible tubers, the external contour of each tuber selected by the machine vision system was fitted with an ellipse using the fitEllipse method from the OpenCV library after edge enhancement based on the Canny algorithm. Upon obtaining the major and minor diameters from the fitted ellipse, a linear regression coefficient was estimated between the actual tuber lengths, measured using a

Vanier calliper, as shown in Figure 4.3, and the estimated tuber lengths derived from the proposed software. This estimation was performed for 20 tubers randomly positioned on the conveyor belt. Equation 4.2 gives the calibration equation between the measured and software-estimated tuber lengths, where l is the millimetre length and n is the number of pixels.

$$l = 0.659\ x\ n \qquad (4.2)$$

(a) (b)

Figure 4.3: Measurement of the (a) major diameter; (b) minor diameter of potato tubers in the Laboratory.

The application of Equation 4.2 had the minimal error in destimating the diameters when the camera was mounted 1100 mm over the converyor, at where the equation was calibrated. Table 1 shows that the error of both diameters increase gradually as the distance between the camera and conveyor drops.

Table 4.1: Size estimates error at different camera heights from the conveyor.

Height above conveyor (mm)	**Absolute minor diameter error (mm)**	**Absolute major diameter error (mm)**
1100.00	1.08	0.16
1050.00	3.49	3.65
1000.00	4.06	9.70
950.00	5.52	13.73

4.6 Experiment for Tuber Size Validation

The size distribution within a representative 22.68 kg bag of potatoes containing 138 tubers was examined to gain insights into the distribution patterns. The tubers were classified into three size classes: big, medium, and small. The results obtained from analyzing a single 50 lb bag are summarized in Table 4.2, showing that the major and minor diameters are directly proportional on average.

An experiment was conducted under two conditions to validate the length estimates from the proposed method: static and moving conveyor belts. The static conveyor experiment aimed to investigate the effect of different clustering scenarios on size estimation accuracy. On the other hand, the moving conveyor experiment aimed to determine whether the motion of the conveyor affects the accuracy of the size estimates. The results of both experiments would provide insights into the applicability of the proposed method in different scenarios where there may be varying degrees of clustering and perspective distortion of potato tubers on the post-harvest conveyors.

Table 4.2: Average minor and major diameters across the different size grades obtained from a single 50 lb bag.

Size grade	Number of tubers	Average minor (mm)	Average major (mm)
Small	112	50	90
Medium	21	60	103
Big	5	68	152

4.6.1 Validation of Size Estimation on Static Conveyor

The potato tubers from three 22.68 kg bags were arranged on the conveyor belt to form the three clustering scenarios, as shown in Figure 3.2. To replicate the perspective distortion observed in real-life scenarios, the tubers were arranged randomly by moving them back and forth on the conveyors shown in Figure 4.2 and ensuring significant

occlusion and clustering were present in the dense scenario. The proposed machine vision system was employed to estimate the sizes of five randomly sampled tubers per frame for 18 frames, with six frames allocated to each clustering scenario. To establish a ground truth for comparison, the 90 selected tubers sized by the proposed system were physically measured using a Vanier calliper. Figure 4.4 shows a densely clustered image in the static conveyor experiment showing the tubers selected for size estimation in yellow bounding boxes.

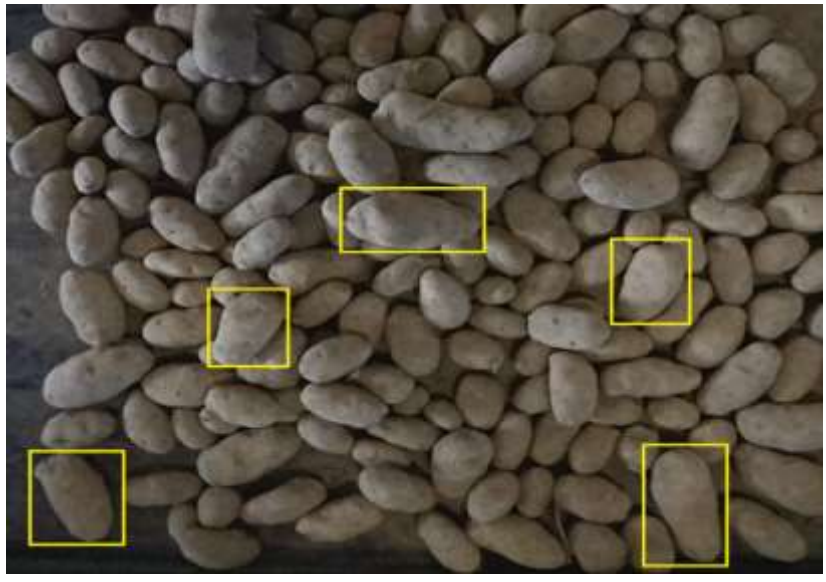


Figure 4.4: Densely occluded image acquired in the static conveyor experiment with the five software-selected tubers in yellow bounding boxes.

4.6.2 Validation of Size Estimation on Moving Conveyor

The proposed method for size estimation of potato tubers utilizes static images captured during conveyance, but it is essential to evaluate the effect of tuber orientation caused by the motion of the conveyor on the accuracy of size estimation. Within this experiment, a batch of 50 tubers was placed onto conveyor (i), shown in Figure 4.2a, which operated at a speed of 0.57 m/s. Conveyors (ii), (iii), and (iv), on the other hand, were set to speeds

of 0.285 m/s, 1.14 m/s, and 0.57 m/s, respectively. By manipulating the conveyor speeds in this manner, it was ensured that the tubers on conveyor (iv) assumed random orientations, thereby enabling the evaluation of the proposed system regarding tuber perspective distortion.

The estimation of the minor and major diameters of the tubers followed a methodology similar to that outlined in Section 4.6.1. However, in contrast to the static conveyor experiment, where five software-selected tubers per frame were utilized, all tubers present in each frame, as shown in Figure 4.5, were used.

Figure 4.5: Image of tubers acquired for the moving conveyor experiment.

4.7 Performance Evaluation

The size estimates obtained from the software for the major and minor diameters of the tubers were assessed using the measured lengths as ground truth. The root mean square error (RMSE), the normalized root mean square error (nRMSE), Lin's Concordance correlation (CCC) and the coefficient of determination (R^2) were evaluated to determine

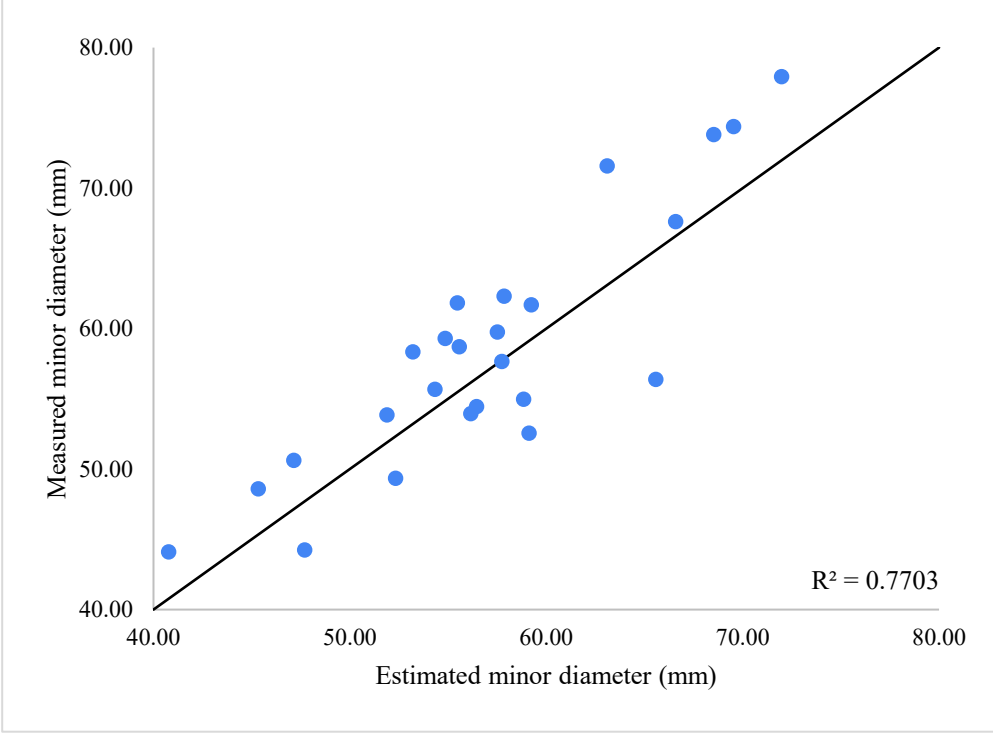
if the software-estimated lengths closely followed the measured lengths of the tubers. Given that the CCC score considers both measurement error and variation, it provides a more comprehensive evaluation of the proposed method, particularly in replacing the existing size estimation system.

4.8 Results of Size Estimation on Static Conveyor

Figures 4.6 – 4.8 show the R^2 for the dense, moderate, and sparse clustering scenarios. The analysis of the results revealed that the proposed system better estimates the major diameter than the minor diameter, irrespective of the clustering scenario. The disparity in the accuracy of measurements can be attributed to the inherent ellipsoidal shape of potato tubers, which gives rise to two minor diameters. To ensure consistency, the maximum of these two minor diameters was used as the ground truth measurement, which may not always be visible to the camera. Moreover, the clustering scenario impacted the estimation of the minor diameter, especially in the dense clustering scenario. This can be attributed to the susceptibility of the minor diameter to perspective distortion, which is further compounded by its smaller size; hence it is more sensitive to variations in the height of the tuber pile, thereby affecting the pixel-to-length calibration given by Equation 4.2. Consequently, the accurate estimation of the minor diameter is subject to additional complexities than the major diameter estimation.

Combining the CCC scores from Table 4.2 and the R^2 scores from Figure 4.6 – 4.8 indicated that the dense clustering performed worst, measuring 0.77 in R^2 and 0.85 in CCC for the minor diameter. These results suggest that the estimated minor diameter explains approximately 77% of the observed variation in the measured minor diameter in the worst-case scenario. Moreover, within the 95% confidence interval of the CCC, from

0.68 to 0.95, the CCC value implies a substantial level of concordance between the measured and estimated diameters.



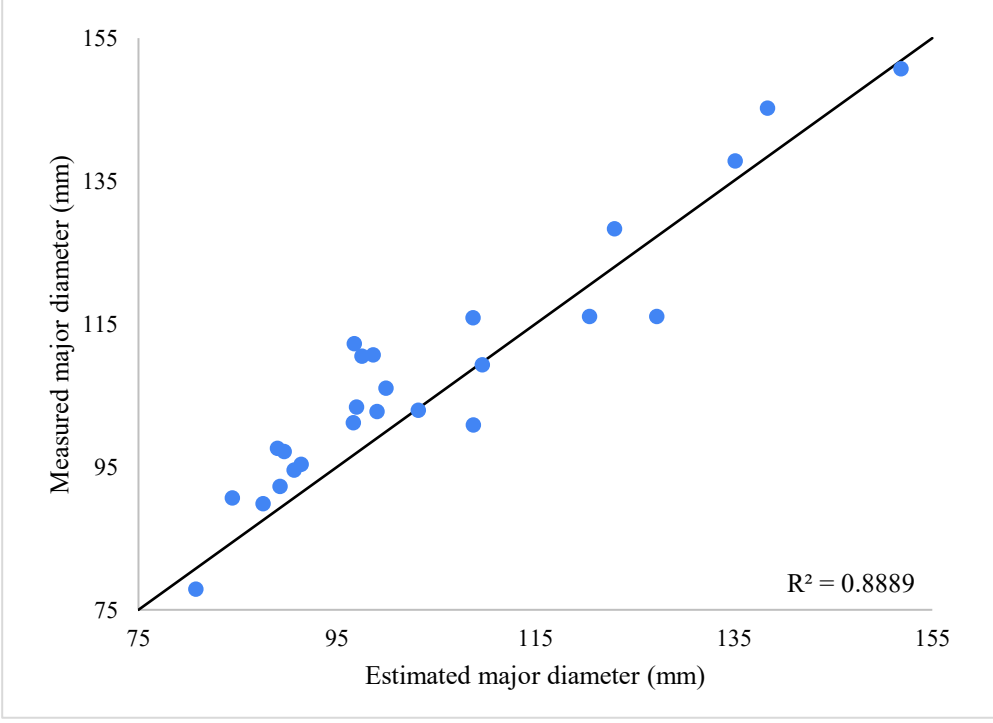
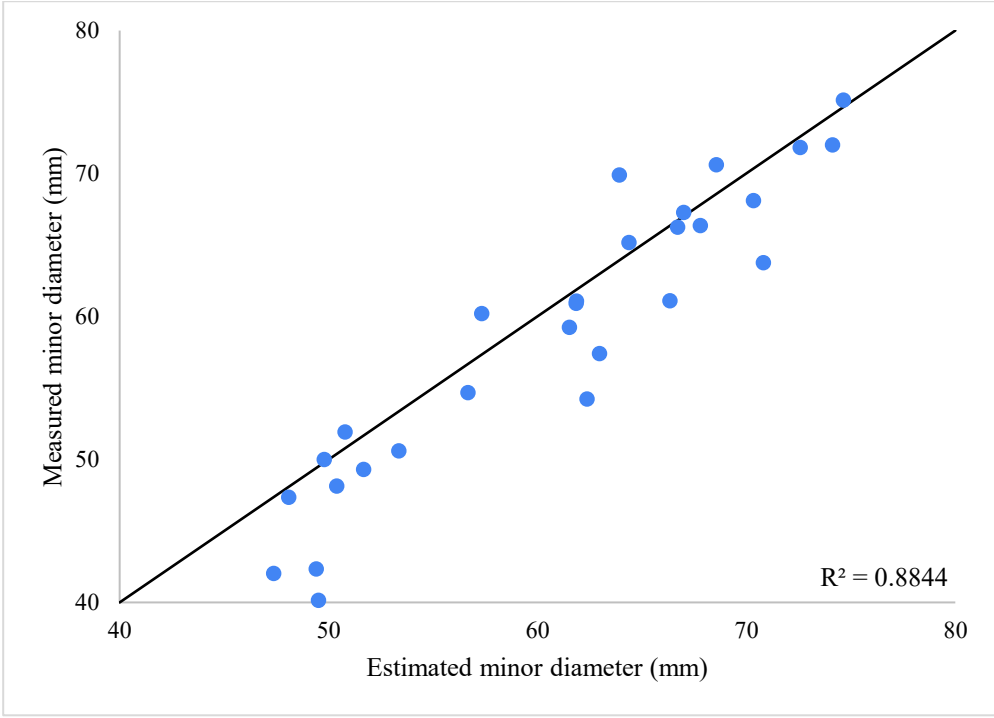


Figure 4.6: Regression analysis of the estimated diameters for dense clustering scenario.



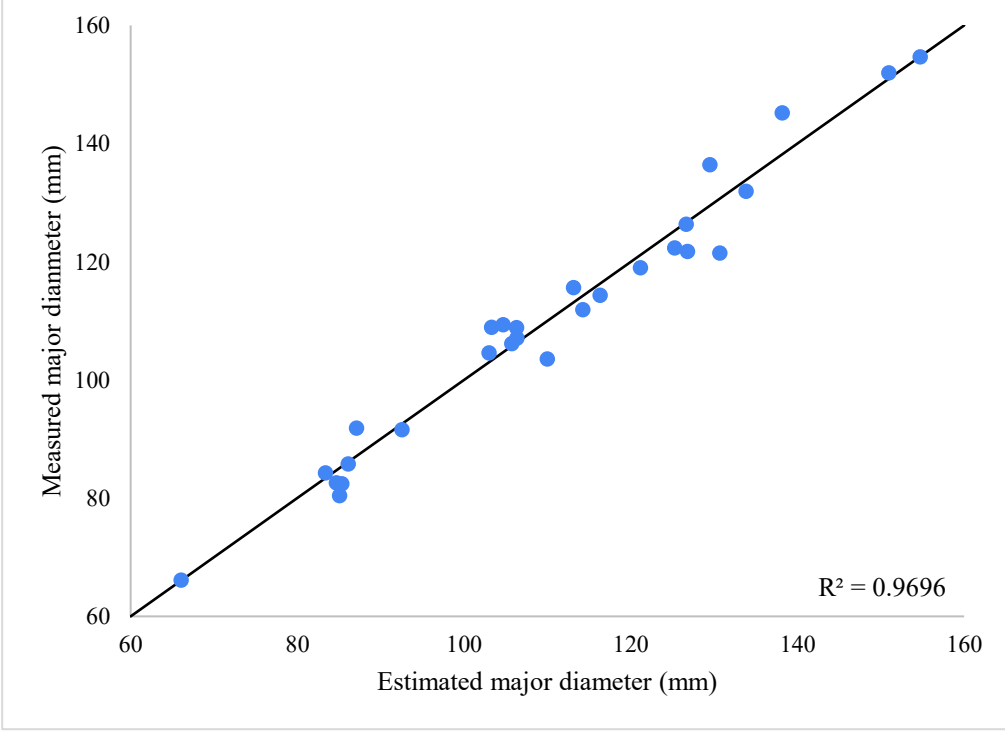
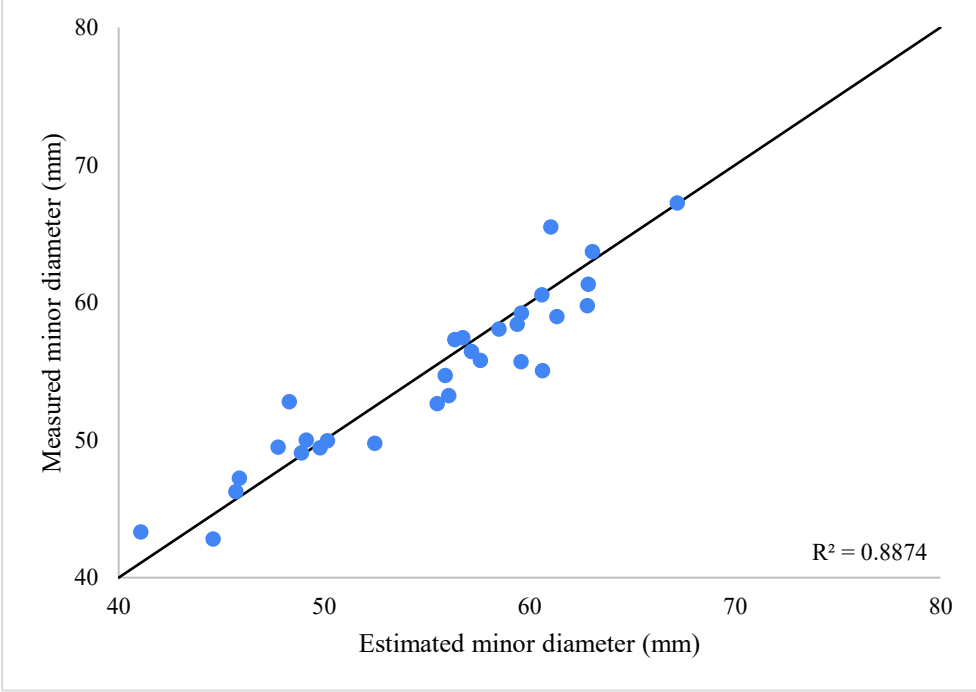


Figure 4.7: Regression analysis of the estimated diameters for moderate clustering scenario.



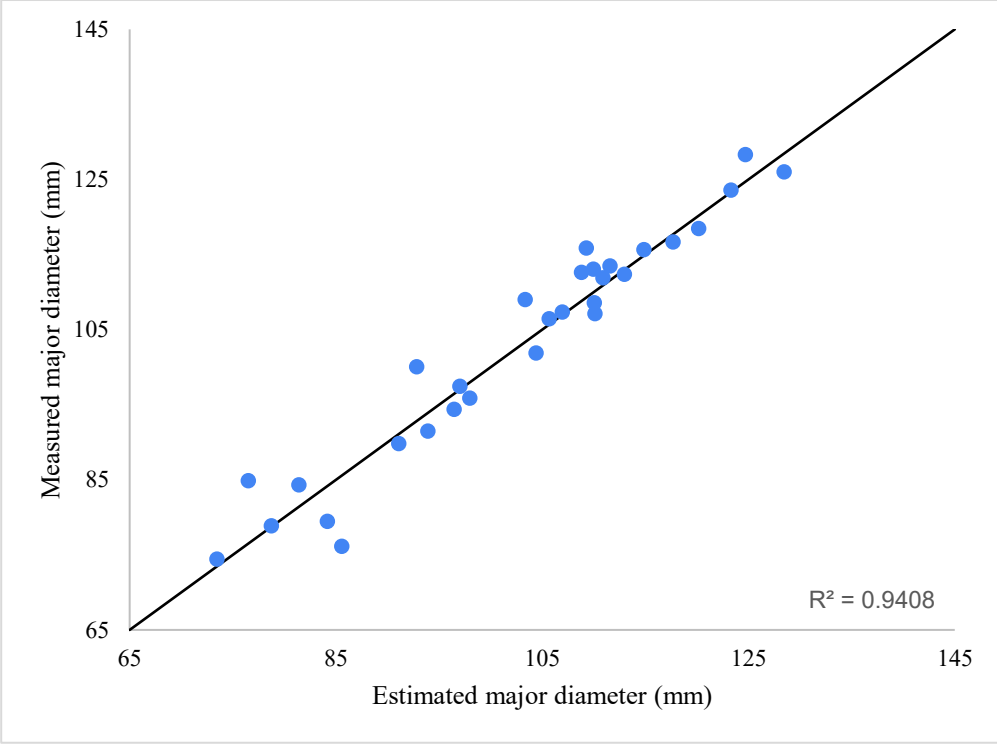


Figure 4.8: Regression analysis of the estimated diameters for sparse clustering scenario.

The CCC score was consistently higher than the R^2 because it considers both deviations from the 45-degree line and the variation (Crawford et al., 2007), whereas the R^2 only measures how much a variation in one variable leads to the variation in another. This suggests that some unexplained variations observed in the measured diameters may be attributed to random errors, which could be mitigated by increasing the number of observations, using a 1:1 pixel camera, and ensuring more consistent ground-truth measurements. However, upon analyzing Figure 4.10, it becomes evident that a partial linear association exists between the residual and estimated minor diameters, exhibiting a Pearson correlation coefficient (r) of 0.47. This finding implies that the errors associated with the minor diameter estimation exhibit a lower degree of randomness when compared to those associated with the major diameter (with r = 0.06). Therefore, introducing a new

feature, such as the distance of the tuber from the camera, can enhance the accuracy of the minor diameter estimates.

Table 4.3: CCC, RMSE, and nRMSE for the estimated lengths based on static conveyor experiment.

Clustering condition	**Minor diameter**			**Major diameter**		
	CCC (-)	RMSE (mm)	nRMSE (%)	CCC (-)	RMSE (mm)	nRMSE (%)
Dense	0.85	4.47	7.83	0.92	7.00	6.70
Moderate	0.91	3.93	6.46	0.98	3.81	3.45
Sparse	0.94	2.23	4.05	0.97	3.68	3.58

Two metrics were utilised to evaluate the error associated with estimating both major and minor diameters: the RMSE and the nRMSE, as shown in Table 4.3. Like the R^2 and CCC results, the dense clustering scenario yielded the highest error for estimating the diameters, with a maximum value of 7.83% for the minor diameter. However, the moderate clustering scenario yielded only about 17.85% improvement in the nRMSE in the minor diameter estimation compared to the dense clustering scenario. This finding suggests that while the clustering scenario primarily influences the estimation of the major diameter, the presence of two minor diameters in potato tubers, which are randomly visible to the camera, further contributes to the error in estimating the minor diameter. Nonetheless, it is significant to note that all errors observed in both minor and major

diameter estimations remained below 10%. This implies that the proposed system maintains an acceptable margin for accurately grading the size of potato tubers.

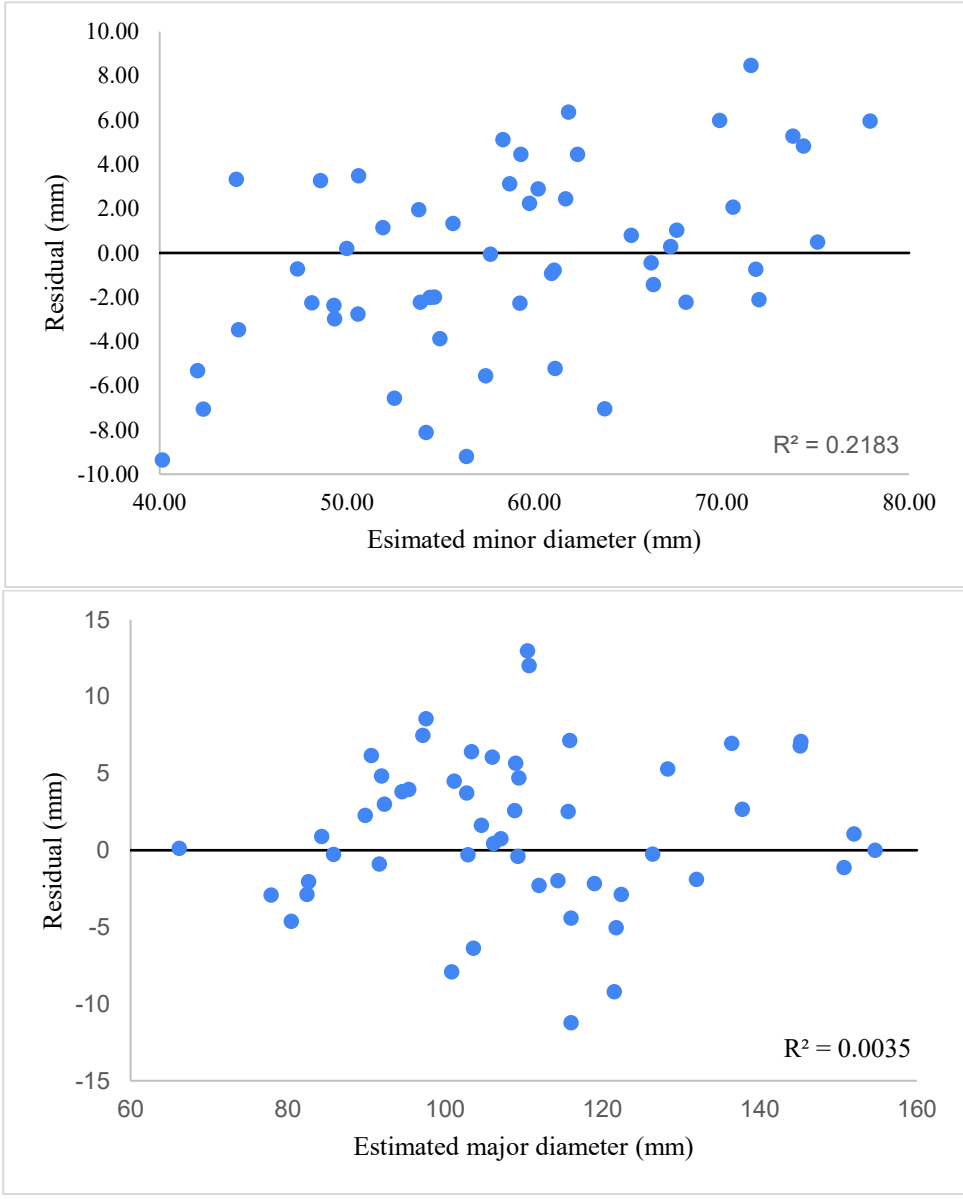


Figure 4.9: Residual scatter plots for the estimated major and minor diameters.

4.9 Results of Size Estimation on Moving Conveyor

The evaluation metrics, including the R^2, CCC, RMSE, and nRMSE scores, are summarized in Table 4.4. In accordance with the findings from the static conveyor experiment, the proposed system demonstrates a higher level of accuracy in estimating

the major diameter of tubers compared to the minor diameter. Moreover, as the total number of tubers used for assessment is increased to 50, notable improvements in R^2 and CCC scores for both the minor and major diameters are observed when compared to the results obtained from the static conveyor experiment. This suggests that the number of sampled tubers influences the accuracy of the measurements, indicating the presence of random errors. Similarly, the nRMSE, which quantifies the deviation in the estimated dimensions, indicates that the proposed system provides a more precise estimation of the major diameter of the tubers. Nevertheless, despite the perspective distortion and varying orientations resulting from running the tubers on multiple conveyors at different speeds, the nRMSE for both the major and minor diameters remains below 10%, which can give a precise size quality representation of the harvested crop.

Table 4.4: Evaluation of the estimated dimensions for the moving conveyor experiment.

Dimension	R^2 (-)	CCC (-)	RMSE (mm)	nRMSE (%)
Minor	0.93	0.96	2.29	4.55
Major	0.98	0.99	1.55	1.60

Considering all clustering scenarios in the moving conveyor experiment was not feasible due to the inherent challenge of precisely tracking a randomly sampled tuber within a moving cluster. However, the findings derived from the static conveyor experiment, in conjunction with the consistently high average mIoU guaranteed by the proposed sampling technique, indicate that the proposed machine vision system can accurately estimate the size of tubers irrespective of the clustering condition and perspective distortion.

4.11 Conclusion

This chapter describes the machine vision system's hardware and the method of evaluating its software performance to estimate the lengths of tubers. Two methods were used to validate the proposed system: static and moving conveyor experiments. The results showed high accuracy in estimating the tuber dimensions when they are free rolling on the conveyors at any orientation and within clusters, as the dimension errors remained below 10% in all scenarios. Estimating the major diameter was better than the minor diameter due to the susceptibility of the minor diameter to perspective distortion. However, increasing the number of tubers for assessing the proposed system reduced the random errors and led to better estimations of the minor and major diameters.

Chapter 5: Conclusion

5.1 On-the-go Deployment of the Machine Vision System in the Field.

The proposed machine vision system was designed with specific constraints to ensure its practicality for use in the field. The primary constraints were related to the system's integration into existing post-harvest conveyors without disrupting farm operations and ensuring system reliability. To achieve seamless integration with farming operations, the proposed system was designed to be autonomous by leveraging the systemd on the Linux OS for automating software execution and using the speed of the conveyor to start/stop image capturing and adjust the camera frame rate. Moreover, the mechanical mount was assembled to fit different conveyors without needing significant modifications to the existing system, as seen in Figure 5.1. Table 5.1 presents an illustrative output obtained from the proposed machine vision system. The Sample ID serves as a unique identifier assigned to each tuber assessed by the system. It is generated by combining the datetime information (year, month, day, hour, minute, and second) corresponding to when the frame was captured with the specific sample number. Consequently, this formulation enables tracking each graded tuber back to a precise moment in time and its corresponding frame. The provided measurements include the length, denoting the major diameter of the tuber, and the width, representing its minor diameter.

Figure 5.1: Test-mounting the machine vision system set-up at McCain Farms of the Future

Table 5.1: Sample output of the proposed machine vision system

Sample ID	Length (mm)	Width (mm)
2023_05_19_08_53_27_01	133.98	73.80
2023_05_19_08_53_27_02	123.75	81.75
2023_05_19_08_53_27_03	138.21	93.14
2023_05_19_08_53_27_04	111.16	76.96
2023_05_19_08_53_27_05	114.93	64.46
2023_05_19_08_53_28_01	101.69	67.72
2023_05_19_08_53_28_02	116.37	65.35
2023_05_19_08_53_28_03	129.19	81.03
2023_05_19_08_53_28_04	122.20	71.55
2023_05_19_08_53_28_05	133.38	79.27

The reliability of machine vision systems can be assessed by evaluating several factors, such as measurement accuracy, system throughput speed, and robustness to varying field conditions. Sections 3.3 and 4.3 present the results related to the accuracy and robustness of the proposed machine vision system, which was above 90% in terms of sampling, segmentation quality, and size measurement accuracy. Moreover, the system throughput was evaluated by measuring the processing time of the computer vision software, which

ranged from image acquisition to size estimation of five tubers per frame, as shown in Table 5.1.

Time-performant software is critical to achieving on-the-go sampling and size estimation. As explained in Section 4.4, to avoid grading the same set of tubers more than once, it was necessary to process images at one frame every two seconds when the conveyor speed was 1.14m/s. However, as seen in Table 5.2, the computer vision software processes one frame in approximately 12 seconds, which could reduce the sample size. Therefore, computational multithreading was employed to address the challenge of acquiring images at a frame rate aligned with the conveyor speed while concurrently processing data read from an image queue in other parts of the software. This ensures the continuous acquisition of images on-the-go which are then continuously processed in the background, guaranteeing the expected sample size by the end of one day of post-harvest operation.

Table 5.2: Time performance-based software analysis using 1 densely occluded image.

Software part	**Execution time (s)**
Image acquisition	0.01
Instance segmentation	6.00
Feature creation	4.30
Sampling	0.01
Write data	1.42
Total	11.74

The current approach for size grading tubers involves using only about 0.1% of the yield and requires human labour. It takes approximately five days to grade 100 hectares of a potato field, as observed at McCain's Farms of the Future. With the proposed machine vision system, the sample size can be increased to approximately 3.5% with a regular sampling frequency. Moreso, utilising a GPU-enabled computer can significantly reduce processing time. This improvement is particularly evident in the performance of the Mask

R-CNN model, which currently requires approximately 6 seconds to detect tuber instances. Additionally, implementing parallel computing techniques can further expedite the process of creating handcrafted features. Consequently, these advancements open up the possibility of achieving real-time sampling and size estimation, providing a substantial leap forward regarding system capabilities.

While the computer vision software was developed and tested on two different varieties of potatoes, generalizing the application for all varieties, which might significantly deviate from an ellipsoid, cannot be verified without testing. Nevertheless, it would be possible to apply techniques such as transfer learning to smoothly increase the scope of varieties detected successfully using this software without the need for complete retraining of the Mask R-CNN model. Recognizing the inherent limitations and taking necessary measures to adapt the software, the machine vision system can be successfully utilized for accurate size estimation across a wide range of potato cultivars. Moreover, with appropriate modifications, there is potential to extend its application to other crops, expanding its usefulness in diverse agricultural contexts.

5.2 Perspectives towards a Comprehensive Non-Destructive Quality Grading of Potato Tubers

The proposed system has potential applications beyond size grading, including shape determination and surface defect detection quality tasks, which can be performed in two dimensions. Additionally, estimating the specific gravity of sampled tubers becomes feasible by integrating depth information and mass flow rate data from load cells. This could be achieved by synchronising load cell-derived mass per frame, size estimates of the sampled tubers obtained from the proposed software, and volume per frame derived from the depth camera, allowing for the calculation of density and subsequent determination of

specific gravity through regression analysis. More so, incorporating a depth camera within the machine vision system enables the calculation of the distance between the potato pile and the camera. This distance can serve as an additional parameter in the calibration equation, enhancing the accuracy of the diameter estimates while eliminating the requirement for different calibration equations for varying camera heights.

One challenge of the current system is the use of RGB cameras, which are restricted to the visible range of the electromagnetic spectrum. This makes using these cameras for non-destructive internal defect grading complex. Replacing RGB cameras with more versatile hyperspectral cameras that can capture a wider range of wavelengths, including the visible near-infrared range, may be necessary to overcome this challenge. This approach can be leveraged for the sampled tubers' chemical and physical quality grading, although it will require careful synchronization of the sampled tubers to their respective spectral and spatial information.

In addition to the quality grading of potato tubers, farmers need to have a comprehensive understanding of the distribution of these quality attributes in the storage facility. This information is critical for efficient planning and logistics. To address this issue, a potato quality map can be developed, which provides a spatial representation of the quality distribution within the storage facilities. This map can help farmers to identify specific locations where tubers with desired quality attributes can be retrieved, leading to better decision-making and efficient use of resources.

5.3 Conclusion

Quality assessment of potato tubers during harvesting is essential to ensure the classification and allocation of potatoes to specific market segments to optimize their utilization and ensure fair compensation for farmers. One such quality assessment carried

out is estimating the tuber size, as this influences the end-use of the tubers. Thus, this research focused on developing a machine vision system for size-grading potato tubers on post-harvest conveyors.

The machine vision system presents practical and compelling prospects for improving the efficiency and precision of size grading operations within the potato industry. Through the automation of tuber size estimation processes in the field, this system can bolster operational efficiency, minimize costs by eliminating reliance on human labour and transportation of tubers for grading, and facilitate improved logistics planning. Additionally, the proposed sampling method has the potential to supplant manual sampling, a procedure that is often prone to size-related biases and susceptibilities arising from human habits.

References

Abbasi, A., Feldman, M. J., Park, J., Greene, K., Novy, R. G., & Liu, H. (2021). A deep learning algorithm for potato tuber hollow heart classification. *BioRxiv*, 2021.09.29.462243. https://doi.org/10.1101/2021.09.29.462243

Abong, G. O., Okoth, M. W., Karuri, E. G., Kabira, J. N., & Mathooko, F. M. (2009). Evaluation of selected Kenyan potato cultivars for processing into French fries. *Journal of Animal & Plant Sciences*, *2*, 141–147. http://www.biosciences.elewa.org/JAPS;

Aggelopoulou, A. D., Bochtis, D., Fountas, S., Swain, K. C., Gemtos, T. A., & Nanos, G. D. (2011). Yield prediction in apple orchards based on image processing. *Precision Agriculture*, *12*(3), 448–456. https://doi.org/10.1007/S11119-010-9187-0/FIGURES/4

Alazab, M., Awajan, A., Mesleh, A., Abraham, A., Jatana, V., & Alhyari, S. (2020). COVID-19 Prediction and Detection Using Deep Learning. *International Journal of Computer Information Systems and Industrial Management Applications*, *12*, 168–181. www.mirlabs.net/ijcisim/index.html

Al-Mallahi, A., Kataoka, T., Okamoto, H., & Shibata, Y. (2010a). An image processing algorithm for detecting in-line potato tubers without singulation. *Computers and Electronics in Agriculture*, *70*(1), 239–244. https://doi.org/10.1016/J.COMPAG.2009.11.001

Al-Mallahi, A., Kataoka, T., Okamoto, H., & Shibata, Y. (2010b). Detection of potato tubers using an ultraviolet imaging-based machine vision system. *Biosystems Engineering*, *105*(2), 257–265. https://doi.org/10.1016/J.BIOSYSTEMSENG.2009.11.004

Andrés, R., Zavala, P., Torriti, M. A. T., & Mmxvii, ©. (2017). *A PATTERN RECOGNITION STRATEGY FOR VISUAL GRAPE BUNCH DETECTION IN VINEYARDS*.

Arshaghi, A., Ghabeli, L., & Ashourin, M. (n.d.). *Detection and Classification of Potato Diseases Using a New Convolution Neural Network Architecture Image Transmission in UAV MIMO UWB-OSTBC System over Rayleigh Channel Using Multiple Description Coding (MDC) View project Detection and Classification of Potato Diseases Potato Using a New Convolution Neural Network Architecture*. https://doi.org/10.18280/ts.380622

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2015). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(12), 2481–2495. https://doi.org/10.48550/arxiv.1511.00561

Bhargava, A., & Bansal, A. (2018). Fruits and Vegetables Quality Evaluation Using Computer Vision: A Review. *Journal of King Saud University - Computer and Information Sciences*, *33*. https://doi.org/10.1016/j.jksuci.2018.06.002

Bhattacharya, S., Reddy Maddikunta, P. K., Pham, Q. V., Gadekallu, T. R., Krishnan S, S. R., Chowdhary, C. L., Alazab, M., & Jalil Piran, M. (2021). Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey. *Sustainable Cities and Society*, *65*, 102589. https://doi.org/10.1016/J.SCS.2020.102589

Binocular stereo vision applied to harvesting robots--《Journal of Jiangsu University(Natural Science Edition)》2008年05期. (n.d.). Retrieved December 9, 2021, from https://en.cnki.com.cn/Article_en/CJFDTotal-JSLG200805002.htm

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

Canziani, A., Paszke, A., & Culurciello, E. (2016). *An Analysis of Deep Neural Network Models for Practical Applications*. https://doi.org/10.48550/arxiv.1605.07678

Chen, L., Jin, S., & Xia, Z. (2021). Towards a Robust Visual Place Recognition in Large-Scale vSLAM Scenarios Based on a Deep Distance Learning. *Sensors 2021, Vol. 21, Page 310*, *21*(1), 310. https://doi.org/10.3390/S21010310

Chen, P., McCarthy, M. J., & Kauten, R. (1989). NMR for internal quality evaluation of fruits and vegetables. *Trans. ASAE*, *32*(5), 1747–1753.

Chen, S. W., Shivakumar, S. S., Dcunha, S., Das, J., Okon, E., Qu, C., Taylor, C. J., & Kumar, V. (2017). Counting Apples and Oranges with Deep Learning: A Data-Driven Approach. *IEEE Robotics and Automation Letters*, *2*(2), 781–788. https://doi.org/10.1109/LRA.2017.2651944

Chicchón, M., & Huerta, R. (2021). Semantic Segmentation Using Convolutional Neural Networks for Volume Estimation of Native Potatoes at High Speed. *Communications in Computer and Information Science*, *1410 CCIS*, 236–249. https://doi.org/10.1007/978-3-030-76228-5_17

Crawford, S. B., Kosinski, A. S., Lin, H. M., Williamson, J. M., & Barnhart, H. X. (2007). Computer programs for the concordance correlation coefficient. *Computer Methods and Programs in Biomedicine*, *88*(1), 62–74. https://doi.org/10.1016/j.cmpb.2007.07.003

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2010). *ImageNet: A large-scale hierarchical image database*. 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Deng, L., Du, H., & Han, Z. (2017). A Carrot Sorting System Using Machine Vision Technique. *Applied Engineering in Agriculture*, *33*(2), 149–156. https://doi.org/10.13031/AEA.11549

Dolata, P., Wróblewski, P., Mrzygłód, M., & Reiner, J. (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*, *190*, 106451. https://doi.org/10.1016/J.COMPAG.2021.106451

Dutta, A., & Zisserman, A. (2019). The VIA Annotation Software for Images, Audio and Video. *Proceedings of the 27th ACM International Conference on Multimedia*. https://doi.org/10.1145/3343031.3350535

Elmasry, G., Cubero, S., Moltó, E., & Blasco, J. (2012). In-line sorting of irregular potatoes by using automated computer-based machine vision system. *Journal of Food Engineering*, *112*(1–2), 60–68. https://doi.org/10.1016/J.JFOODENG.2012.03.027

FAO publications catalogue 2021. (2021). *FAO Publications Catalogue 2021*. https://doi.org/10.4060/CB4402EN

Farhadi, R., & Ghanbarian, D. (2014). POTATO MASS MODELING WITH DIMENSIONAL ATTRIBUTES USING REGRESSION AND ARTIFICIAL NEURAL NETWORKS. *Trakia Journal of Sciences*, *12*, 47–54.

Faria, F. T. J., Bin Moin, M., Al Wase, A., Sani, M. R., Hasib, K. M., & Alam, M. S. (2023). Classification of Potato Disease with Digital Image Processing Technique: A Hybrid Deep Learning Framework. *2023 IEEE 13th Annual Computing and Communication Workshop and Conference, CCWC 2023*, 820–826. https://doi.org/10.1109/CCWC57344.2023.10099162

Firouzjaei, R. A., Minaei, S., & Beheshti, B. (2018). Sweet lemon mechanical damage detection using image processing technique and UV radiation. *Journal of Food Measurement and Characterization*, *12*(3), 1513–1518. https://doi.org/10.1007/S11694-018-9766-8/FIGURES/5

Fu, L., Tola, E., Al-Mallahi, A., Li, R., & Cui, Y. (2019). A novel image processing algorithm to separate linearly clustered kiwifruits. *Biosystems Engineering*, *183*, 184–195. https://doi.org/10.1016/J.BIOSYSTEMSENG.2019.04.024

Gadekallu, T. R., Khare, N., Bhattacharya, S., Singh, S., Maddikunta, P. K. R., & Srivastava, G. (2020). Deep neural networks to predict diabetic retinopathy. *Journal of Ambient Intelligence and Humanized Computing*, *1*, 1–14. https://doi.org/10.1007/S12652-020-01963-7/TABLES/4

Garcia-Lamont, F., Cervantes, J., López, A., & Rodriguez, L. (2018). Segmentation of images by color features: A survey. *Neurocomputing*, *292*, 1–27. https://doi.org/10.1016/J.NEUCOM.2018.01.091

Girshick, R. (2015). *Fast R-CNN* (pp. 1440–1448).

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2013). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 580–587. https://doi.org/10.48550/arxiv.1311.2524

Hassankhani, R., & Navid, H. (2012). Potato Sorting Based on Size and Color in Machine Vision System An intelligent system for assessing the quality of the cereal sowing. View project Application of satellite image in landuse map extraction View project Potato Sorting Based on Size and Color in Machine Vision System. *Article in Journal of Agricultural Science*, *4*(5). https://doi.org/10.5539/jas.v4n5p235

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 386–397. https://doi.org/10.48550/arxiv.1703.06870

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). *Mask R-CNN* (pp. 2961–2969).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep Residual Learning for Image Recognition* (pp. 770–778).

Heikamp, K., & Bajorath, J. (2013). Support vector machines for drug discovery. *Http://Dx.Doi.Org/10.1517/17460441.2014.866943*, *9*(1), 93–104. https://doi.org/10.1517/17460441.2014.866943

Huda, N., Sarker, K. U., & Munnaf, M. A. (2019). Design, fabrication and performance evaluation of a small drum type potato grader. *CIGR JOURNAL*, *21*(4), 107–114. http://hdl.handle.net/1854/LU-8709538

Ishimwe, R., Abutaleb, K., & Ahmed, F. (2014). Applications of Thermal Imaging in Agriculture—A Review. *Advances in Remote Sensing*, *03*(03), 128–140. https://doi.org/10.4236/ARS.2014.33011

Ismail, N., & Malik, O. A. (2022). Real-time visual inspection system for grading fruits using computer vision and deep learning techniques. *Information Processing in Agriculture*, *9*(1), 24–37. https://doi.org/10.1016/J.INPA.2021.01.005

Jang, S. H., Moon, S. P., Kim, Y. J., & Lee, S. H. (2023). Development of Potato Mass Estimation System Based on Deep Learning. *Applied Sciences 2023, Vol. 13, Page 2614*, *13*(4), 2614. https://doi.org/10.3390/APP13042614

Ji, B., Zhu, W., Liu, B., Ma, C., & Li, X. (2009). Review of recent machine-vision technologies in agriculture. *2009 2nd International Symposium on Knowledge Acquisition and Modeling, KAM 2009*, *3*, 330–334. https://doi.org/10.1109/KAM.2009.231

Kabira, J. N., & Lemaga, B. (n.d.). *Potato processin quality evaluation procedures for research and food industry applications in East and Central Africa*.

Kandimalla, V. V. (2021). *DEEP LEARNING APPROACHES TO CLASSIFY AND TRACK AT-RISK FISH SPECIES*.

Kataoka, T., Kaneko, T., Okamoto, H., & Hata, S. (2003). Crop growth estimation system using machine vision. *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM*, *2*, 1079–1083. https://doi.org/10.1109/AIM.2003.1225492

Koirala, A., Walsh, K. B., Wang, Z., & McCarthy, C. (2019). Deep learning – Method overview and review of use for fruit detection and yield estimation. *Computers and Electronics in Agriculture*, *162*, 219–234. https://doi.org/10.1016/J.COMPAG.2019.04.017

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (n.d.). *ImageNet Classification with Deep Convolutional Neural Networks*. Retrieved April 5, 2022, from http://code.google.com/p/cuda-convnet/

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature 2015 521:7553*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, H. S., & Shin, B. S. (2020). Potato Detection and Segmentation Based on Mask R-CNN. *Journal of Biosystems Engineering*, *45*(4), 233–238. https://doi.org/10.1007/S42853-020-00063-W/METRICS

Lee, Y. J., & Shin, B. S. (2020). Development of Potato Yield Monitoring System Using Machine Vision. *Journal of Biosystems Engineering*, *45*(4), 282–290. https://doi.org/10.1007/S42853-020-00069-4/METRICS

Lili, W., Bo, Z., Jinwei, F., Xiaoan, H., Shu, W., Yashuo, L., Qiangbing, Z., Chongfeng, W., L, W. L., W, F. J., & A, H. X. (2017). Development of a tomato harvesting robot used in greenhouse. *International Journal of Agricultural and Biological Engineering*, *10*(4), 140–149. https://doi.org/10.25165/IJABE.V10I4.3204

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8693 LNCS*(PART 5), 740–755. https://doi.org/10.48550/arxiv.1405.0312

Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). *Feature Pyramid Networks for Object Detection* (pp. 2117–2125).

Liu, J., & Wang, X. (2021). Plant diseases and pests detection based on deep learning: a review. *Plant Methods*, *17*(1), 1–18. https://doi.org/10.1186/S13007-021-00722-9/TABLES/4

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2015). SSD: Single Shot MultiBox Detector. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9905 LNCS*, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Mahajan, S., Das, A., & Sardana, H. K. (2015). Image acquisition techniques for assessment of legume quality. *Trends in Food Science & Technology*, *42*(2), 116–133. https://doi.org/10.1016/J.TIFS.2015.01.001

Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0B013E3181EC173D

Marino, S., Smolarz, A., & Beauseroy, P. (2019). Potato defects classification and localization with convolutional neural networks. *Https://Doi.Org/10.1117/12.2521264*, *11172*, 110–117. https://doi.org/10.1117/12.2521264

Marwaha, R. S., Pandey, S. K., Kumar, D., Singh, S. V., & Kumar, P. (2010). Potato processing scenario in India: Industrial constraints, future projections, challenges ahead and remedies - A review. *Journal of Food Science and Technology*, *47*(2), 137–156. https://doi.org/10.1007/S13197-010-0026-0/METRICS

Mavridou, E., Vrochidou, E., Papakostas, G. A., Pachidis, T., & Kaburlasos, V. G. (2019). Machine Vision Systems in Precision Agriculture for Crop Farming. *Journal of Imaging 2019, Vol. 5, Page 89*, *5*(12), 89. https://doi.org/10.3390/JIMAGING5120089

McCarthy, M. J., & McCarthy, K. L. (1994). Quantifying transport phenomena in food processing with nuclear magnetic resonance imaging. *Journal of the Science of Food and Agriculture*, *65*(3), 257–270. https://doi.org/10.1002/JSFA.2740650302

Mechanical potato graders. (2020). https://blog.potatoworld.eu/mechanical-potato-graders

Mirbod, O., Choi, D., Heinemann, P., & Marini, R. (2020). *Towards Image-Based Measurement of Accurate Apple Size and Yield Using Stereo Vision Cameras*. *ASABE 2020 Annual International Meeting*, 1-. https://doi.org/10.13031/AIM.202001115

Narvankar, D. S., Jha, S. K., & Singh, A. (2005). Development of Rotating Screen Grader for selected orchard crops. *Journal of Agricultural Engineering*, *42*(4), 60–64. https://www.indianjournals.com/ijor.aspx?target=ijor:joae&volume=42&issue=4&article=012

Noordam, J. C., Otten, G. W., Timmermans, T. J. M., & Zwol, B. H. van. (2000). High-speed potato grading and quality inspection based on a color vision system. *Https://Doi.Org/10.1117/12.380075*, *3966*, 206–217. https://doi.org/10.1117/12.380075

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Pandey, N., Kumar, S., & Pandey, R. (2019). Grading and defect detection in potatoes using deep learning. *Communications in Computer and Information Science*, *839*, 329–339. https://doi.org/10.1007/978-981-13-2372-0_29/FIGURES/6

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32* (pp. 8024–8035). Curran Associates, Inc. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Patel, K. K., Kar, A., Jha, S. N., & Khan, M. A. (2012). Machine vision system: A tool for quality inspection of food and agricultural products. *Journal of Food Science and Technology*, *49*(2), 123–141. https://doi.org/10.1007/S13197-011-0321-4/TABLES/4

Pavlista, A. D., & Ojala, J. C. (1997). Potatoes: Chip and French Fry Processing. *Processing Vegetables*, 237–284. https://doi.org/10.1201/9780203741863-12/POTATOES-CHIP-FRENCH-FRY-PROCESSING-ALEXANDER-PAVLISTA-JOHN-OJALA

Plá, F., Juste, F., & Ferri, F. (1993). Feature extraction of spherical objects in image analysis: an application to robotic citrus harvesting. *Computers and Electronics in Agriculture*, *8*(1), 57–72. https://doi.org/10.1016/0168-1699(93)90058-9

Potato Facts and Figures - International Potato Center. (2021). https://cipotato.org/potato/potato-facts-and-figures/

Potato market information review, 2020-2021 - agriculture.canada.ca. (n.d.). Retrieved May 10, 2022, from https://agriculture.canada.ca/en/canadas-agriculture-sectors/horticulture/horticulture-sector-reports/potato-market-information-review-2020-2021

Potato production in Canada – Canadian Horticultural Council. (2018). https://hortcouncil.ca/about-us/horticulture-stats/potato-production-in-canada/

Prasetyo, N. A., Pranowo, & Santoso, A. J. (2020). Automatic Detection and Calculation of Palm Oil Fresh Fruit Bunches using Faster R-CNN. *International Journal of Applied Science and Engineering*, *17*(2), 121–134. https://doi.org/10.6703/IJASE.202005_17(2).121

Quilloy, E. P., & Bato, P. M. (2015). Machine vision-based software for automating the grading process of Philippine table eggs. *Philippine Agricultural Scientist*, *98*(2), 148–156.

Rady, A. M., & Guyer, D. E. (2015). Rapid and/or nondestructive quality evaluation methods for potatoes: A review. *Computers and Electronics in Agriculture*, *117*, 31–48. https://doi.org/10.1016/J.COMPAG.2015.07.002

Razmjooy, N., Mousavi, B. S., & Soleymani, F. (2012). A real-time mathematical computer method for potato inspection using machine vision. *Computers & Mathematics with Applications*, *63*(1), 268–279. https://doi.org/10.1016/J.CAMWA.2011.11.019

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-December*, 779–788. https://doi.org/10.48550/arxiv.1506.02640

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems*, *28*. https://github.com/

Rigney, M. P., Brusewitz, G. H., & Stone, M. L. (1996). Peach physical characteristics for orientation. *Transactions of the American Society of Agricultural Engineers*, *39*(4), 1493–1497. https://doi.org/10.13031/2013.27643

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9351*, 234–241. https://doi.org/10.48550/arxiv.1505.04597

Samatha, B., Rao, D. K., Syamsundararao, T., Mani, G., Karyemsetty, N., & Santhi, M. V. B. T. (2023). Classification of potato diseases using deep learning apprach. *Proceedings of the International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics, ICIITCEE 2023*, 508–513. https://doi.org/10.1109/IITCEE57236.2023.10090983

Sanchez, P. D. C., Hashim, N., Shamsudin, R., & Mohd Nor, M. Z. (2020). Applications of imaging and spectroscopy techniques for non-destructive quality evaluation of potatoes and sweet potatoes: A review. *Trends in Food Science & Technology*, *96*, 208–221. https://doi.org/10.1016/J.TIFS.2019.12.027

Sannakki, S. S., Rajpurohit, V. S., Nargund, V. B., & Kulkarni, P. (2013). Diagnosis and classification of grape leaf diseases using neural networks. *2013 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013*. https://doi.org/10.1109/ICCCNT.2013.6726616

Sannakki, S. S., Rajpurohit, V. S., Nargund, V. B., Kumar, A., & Yallur, P. S. (2011). Leaf disease grading by machine vision and fuzzy logic. *Int J*, *2*(5), 1709–1716.

Sethy, P. K., Routray, B., & Behera, S. K. (2019). Detection and Counting of Marigold Flower Using Image Processing Technique. *Lecture Notes in Networks and Systems*, *41*, 87–93. https://doi.org/10.1007/978-981-13-3122-0_9

Shahin, M. A., & Tollner, E. W. (1997). Apple classification based on watercore features using fuzzy logic. *Paper American Society of Agricultural Engineers*, *1*, 973–977.

Shelhamer, E., Long, J., & Darrell, T. (2014). Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(4), 640–651. https://doi.org/10.48550/arxiv.1411.4038

Shen, D., Zhang, S., Ming, W., He, W., Zhang, G., & Xie, Z. (2022). Development of a new machine vision algorithm to estimate potato's shape and size based on support vector machine. *Journal of Food Process Engineering*, *45*(3), e13974. https://doi.org/10.1111/JFPE.13974

Shinde, P. P., & Shah, S. (2018). A Review of Machine Learning and Deep Learning Applications. *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018*. https://doi.org/10.1109/ICCUBEA.2018.8697857

Si, Y., Sankaran, S., Knowles, N. R., & Pavek, M. J. (2017). Potato Tuber Length-Width Ratio Assessment Using Image Analysis. *American Journal of Potato Research*, *94*(1), 88–93. https://doi.org/10.1007/S12230-016-9545-1/TABLES/2

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. https://doi.org/10.48550/arxiv.1409.1556

Sites, P. W., & Delwiche, M. J. (1988). Computer Vision to Locate Fruit on a Tree. *Transactions of the ASAE*, *31*(1), 257–0265. https://doi.org/10.13031/2013.30697

Sivaranjani, A., Senthilrani, S., kumar, B. A., & Murugan, A. S. (2021). An Overview of Various Computer Vision-based Grading System for Various Agricultural Products. *Https://Doi.Org/10.1080/14620316.2021.1970631*, 1–23. https://doi.org/10.1080/14620316.2021.1970631

Smith, L. N., Zhang, W., Hansen, M. F., Hales, I. J., & Smith, M. L. (2018). Innovative 3D and 2D machine vision methods for analysis of plants and crops in the field. *Computers in Industry*, *97*, 122–131. https://doi.org/10.1016/J.COMPIND.2018.02.002

Stoll, M., & Jones, H. G. (2007). Thermal imaging as a viable tool for monitoring plant stress. *OENO One*, *41*(2), 77–84. https://doi.org/10.20870/OENO-ONE.2007.41.2.851

Su, Q., Kondo, N., Al Riza, D. F., & Habaragamuwa, H. (2020). Potato quality grading based on depth imaging and convolutional neural network. *Journal of Food Quality*, *2020*. https://doi.org/10.1155/2020/8815896

Su, Q., Kondo, N., al Riza, D. F., & Habaragamuwa, H. (2020). Potato quality grading based on depth imaging and convolutional neural network. *Journal of Food Quality*, *2020*. https://doi.org/10.1155/2020/8815896

Su, Q., Kondo, N., Li, M., Sun, H., & Al Riza, D. F. (2017). Potato feature prediction based on machine vision and 3D model rebuilding. *Computers and Electronics in Agriculture*, *137*, 41–51. https://doi.org/10.1016/J.COMPAG.2017.03.020

Su, Q., Kondo, N., Li, M., Sun, H., al Riza, D. F., & Habaragamuwa, H. (2018). Potato quality grading based on machine vision and 3D shape analysis. *Computers and Electronics in Agriculture*, *152*, 261–268. https://doi.org/10.1016/J.COMPAG.2018.07.012

Su, Q., Kondo, N., Li, M., Sun, H., Al Riza, D. F., & Habaragamuwa, H. (2018). Potato quality grading based on machine vision and 3D shape analysis. *Computers and Electronics in Agriculture*, *152*, 261–268. https://doi.org/10.1016/J.COMPAG.2018.07.012

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2016). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 4278–4284. https://doi.org/10.48550/arxiv.1602.07261

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). Going Deeper with Convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June-2015*, 1–9. https://doi.org/10.48550/arxiv.1409.4842

Tabatabaeefar, A. (2002). Size and shape of potato tubers. *International Agrophysics*, *16*.

Tang, L., Tian, L., & Steward, B. L. (2000). COLOR IMAGE SEGMENTATION WITH GENETIC ALGORITHM FOR IN-FIELD WEED SENSING. *Transactions of the ASAE*, *43*(4), 1019-. https://doi.org/10.13031/2013.2970

Throop, J. A., Aneshansley, D. J., & Upchurch, B. L. (1993). Near-IR and color imaging for bruise detection on Golden Delicious apples. *Https://Doi.Org/10.1117/12.144042*, *1836*, 33–44. https://doi.org/10.1117/12.144042

Tian, H., Wang, T., Liu, Y., Qiao, X., & Li, Y. (2020). Computer vision technology in agricultural automation —A review. *Information Processing in Agriculture*, *7*(1), 1–19. https://doi.org/10.1016/J.INPA.2019.09.006

Valentin, M. T., Villota, E. M., Malamug, V. U., & Agulto, I. C. (2016). Evaluation of a Helix Type Potato Grader. *CLSU International Journal of Science & Technology (2016)*, *1*(1), 46–52. https://doi.org/10.22137/IJST.2016.V1N1.05

Verma, S. R., & Kalkat, H. S. (1975). Design and development of an expanding pitch rubber spool potato sizer. *Journal of Agricultural Engineering*.

Visually estimating potato size grades. (2015). https://blog.potatoworld.eu/visually-estimating-potato-size-grades

Wan, P., Toudeshki, A., Tan, H., & Ehsani, R. (2018). A methodology for fresh tomato maturity detection using computer vision. *Computers and Electronics in Agriculture*, *146*, 43–50. https://doi.org/10.1016/J.COMPAG.2018.01.011

Wang, C., & Xiao, Z. (2021). Potato Surface Defect Detection Based on Deep Transfer Learning. *Agriculture 2021, Vol. 11, Page 863*, *11*(9), 863. https://doi.org/10.3390/AGRICULTURE11090863

Wang, Q., Nuske, S., Bergerman, M., & Singh, S. (2013). Automated Crop Yield Estimation for Apple Orchards. *STAR*, *88*, 745–758. https://doi.org/10.1007/978-3-319-00065-7_50

Wenhua, M., Baoping, J., Jicheng, Z., Xiaochao, Z., & Xiaoan, H. (2009). Apple location method for the apple harvesting robot. *Proceedings of the 2009 2nd International Congress on Image and Signal Processing, CISP'09*. https://doi.org/10.1109/CISP.2009.5305224

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019). *Detectron2*.

Xiong, X. M., Hua, C. J., Fang, C. J., & Chen, Y. (2016). Research on free-form surface stereo matching method based on improved Census transform. *Chinese Control Conference, CCC*, *2016-August*, 4889–4893. https://doi.org/10.1109/CHICC.2016.7554112

Yang, B., & Xu, Y. (2021). Applications of deep-learning approaches in horticultural research: a review. *Horticulture Research*, *8*(1), 123. https://doi.org/10.1038/S41438-021-00560-9/42044846/41438_2021_ARTICLE_560.PDF

Yimyam, P. T. P. S. (2005). Physical Properties Analysis of Mango using Computer Vision. 제어로봇시스템학회:학술대회논문집, 746–750.

Yud-RenChen, K. C., & Kim, M. S. (2002). *Machine vision technology for agricultural applications*. Elsevier Computers and Electronics in Agriculture (36 2002) 173_191.

Zarzecka, K., Gugała, M., Sikorska, A., Grzywacz, K., & Niewęgłowski, M. (2020). Marketable Yield of Potato and Its Quantitative Parameters after Application of Herbicides and Biostimulants. *Agriculture 2020, Vol. 10, Page 49*, *10*(2), 49. https://doi.org/10.3390/AGRICULTURE10020049

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1253. https://doi.org/10.1002/WIDM.1253

Zhao, J., Liu, J., Chen, Q., & Vittayapadung, S. (2008). Detecting subtle bruises on fruits with hyperspectral imaging. *Nongye Jixie Xuebao/Transactions of the Chinese Society of Agricultural Machinery*, *39*, 106–109.

Appendix A: Additional Tables

Table A-1: Matrix for Image Collection at McCain Farm of the Future

	Sunny	**Shaded**	**Fluctuating light**	**Cloudy**
Morning	60 mins	60 mins	15 mins	15 mins
Afternoon	60 mins	60 mins	15 mins	15 mins
Evening	60 mins	60 mins	15 mins	15 mins

Table A-2: Parameter name, value, and description of each hyper-parameter used for the Mask R-CNN model.

Parameter Name	**Value**	**Description**
BACKBONE	ResNet-101 ResNet-50	The convolutional neural network used in the first stage of the Mask R-CNN. Two (i.e., mask_rcnn_R_50_FPN_3x and mask_rcnn_50_C4_3x) of the three pre-trained weights were based on ResNet-50.
NUM_WORKERS	2	Number of GPUs used for training the model. A higher number can improve the training speed.
IMS_PER_BATCH	2	The number of images fed per batch during training. A higher value corresponds to higher memory usage.
MAX_ITER	500	The maximum number of training iterations.
WARMUP_ITERS	1000	The total number of warm-up iterations at the beginning of training.
STEPS	1000, 1500	The iteration number to decrease the learning rate by GAMMA.
GAMMA	0.5	Determines the factor by which to decrease the learning rate at each step. A larger value means that the learning rate will be reduced more quickly, which can lead to a suboptimal solution.
BASE_LR	0.00025	The learning rate for every group by default. A larger value can result in early convergence.
BATCH_SIZE_PER_IMAGE	128	The number of regions per image used to train the RPN.
NUM_CLASSES	1	Number of foreground classes
EVAL_PERIOD	500	The period (in terms of steps) to evaluate the model during training.

Table A-3: Pseudo-code of the image pre-processing steps for creating the avg_val_bbox

Algorithm: Image pre-processing for creating avg_val_bbox

Input: path of image dataset of potato tubers

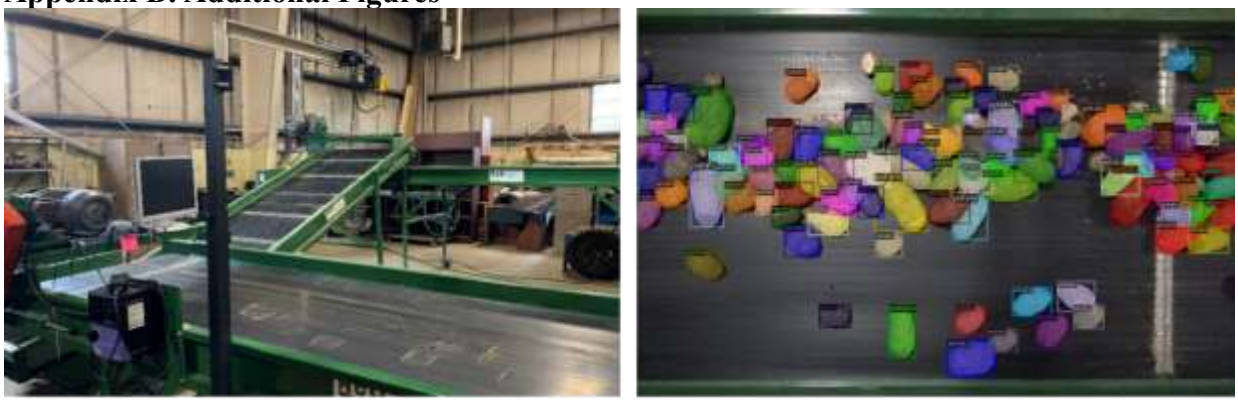
Step 1: initialize variables
x ← [] {x coordinates of the detected bbox}
y ← [] {y co-ordinates of the detected bbox}
arr1 ← [] {array of 1s and 0s for the bottom edge of the bbox}
arr2 ← [] {array of 1s and 0s for the top edge of the bbox}
arr3 ← [] {array of 1s and 0s for the left edge of the bbox}
arr4 ← [] {array of 1s and 0s for the right edge of the bbox}
Step 2: load data
Image ← image; {input image with grayscale}
cv ← openCV2; {python library for computer vision task}
Step 3: image binarization
Image ← cv.threshold (Image, threshold); {image binarization}
Step 4: find contours and obtain the detected bbox co-ordinates from the pre-processed image.
contours ← cv.findContours; {contour consists of four co-ordinates of the bbox}
for each contour in contours do
for each i and j value in contour do
x ← x.append(i);
y ← y.append(j);
Step 5: find the mask pixels on the detected bbox edges
for each i in range(min(x), max(x)+1) {scan through the horizontal bbox edges}
if Image[min(y) + 1, i] == 255 {1 pixel from the bottom edge if pixel value is 255}
arr1 ← arr1.append(1);
else
arr1 ← arr1.append(0);
if Image[max(y) - 1, i] == 255 {1 pixel from the top edge if pixel value is 255}
arr2 ← arr2.append(1);
else
arr2 ← arr2.append(0);
for each i in range(min(y), max(y)+1) {scan through the vertical bbox edges}
if Image[i, min(x) + 1] == 255 {1 pixel from the left edge if pixel value is 255}
arr3 ← arr3.append(1);
else
arr1 ← arr1.append(0);

Algorithm: Image pre-processing for creating avg_val_bbox
if Image[i, max(x) - 1] == 255 {1 pixel from the right edge if pixel value is 255} arr4 ← arr4.append(1); else arr4 ← arr4.append(0); Step 6: find the average pixel count on detected bbox edges avg_val_bbox ← mean(mean(arr1), mean(arr2), mean(arr3), mean(arr4));

Table A-4: Computer specifications (Nuvo 7006E)

Attribute	**Value**
Brand Name	Nuvo
Series	7000E/7000DE/7000P
Model Number	7006E
Item Dimensions	240 mm (W) x 225 mm (D) x 90 mm (H)
Item Weight	3.58 kg
Operating Temperature	-25°C/+70°C
Processor Type	Intel® 8th-Gen Coffee Lake Core™
Number of Processors	6
Memory Size	32 GB
Memory Type	DDR4
Hard Disk Size	500 GB
Hard Disk Interface	Solid State
Operating System	Linux Ubuntu 22.04

Appendix B: Additional Figures



(a) (b)

Figure B-1: (a) Laboratory set-up for image acquisition and (b) annotated image sample of tubers on the laboratory conveyor (section 3.2.2)

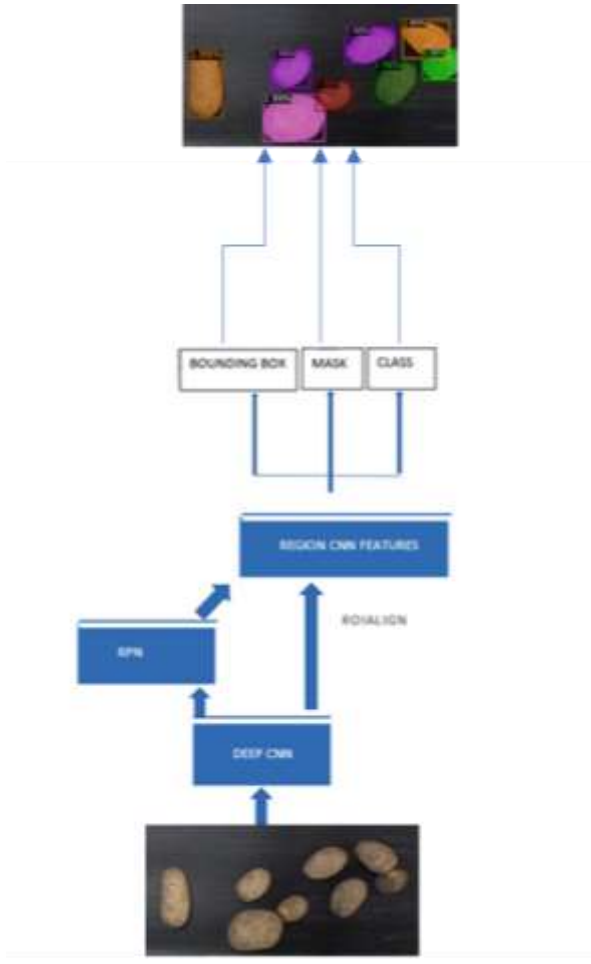


Figure B-2: Architecture of Mask R-CNN for instance segmentation (Kandimalla, 2021)

Appendix C: Hu Moments

The invariant image moments of an image of order (p + q) can be achieved using central moments, which are defined as follows:

$$mu_{pq} = \frac{\sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q I(x,y)}{\left(\sum_x \sum_y I(x,y)\right)^{(p+q)/2+1}} \quad \text{(C-1)}$$

where $I(x, y)$ is the intensity of the pixel at location (x, y) in the image, and $\bar{x}$ and $\bar{y}$ are the coordinates of the centroid of the image, which are given by:

$$\bar{x} = \frac{\sum_x \sum_y xI(x,y)}{\sum_x \sum_y I(x,y)} \quad \text{(C-2)}$$

$$\bar{y} = \frac{\sum_x \sum_y yI(x,y)}{\sum_x \sum_y I(x,y)} \quad \text{(C-3)}$$

The seven Hu moments are obtained using the following formulas:

$$h1 = (mu20 + mu02) \quad \text{(C-4)}$$

$$h2 = ((mu20 - mu02)^2 + 4 * mu11^2)^{0.5} \quad \text{(C-5)}$$

$$h3 = ((mu30 - 3mu12)^2 + (3mu21 - mu03)^2)^{0.5} \quad \text{(C-6)}$$

$$h4 = ((mu30 + mu12)^2 + (mu21 + mu03)^2)^{0.5} \quad \text{(C-7)}$$

$$h5 = (mu30 - 3mu12)(mu30 + mu12)((mu30 + mu12)^2 - 3(mu21 + mu03)^2) + (3mu21 - mu03)(mu21 + mu03)(3(mu30 + mu12)^2 - (mu21 + mu03)^2) \quad \text{(C-8)}$$

$$h6 = (mu20 - mu02)((mu30 + mu12)^2 - (mu21 + mu03)^2) + 4mu11 * (mu30 + mu12) * (mu21 + mu03) \quad \text{(C-9)}$$

$$h7 = (3mu21 - mu03)(mu30 + mu12)((mu30 + mu12)^2 - 3(mu21 + mu03)^2) -$$

$$(mu30 - 3mu12)(mu21 + mu03)(3(mu30 + mu12)^2 - (mu21 + mu03)^2) \quad \text{(C-10)}$$

Appendix D: Evaluation metrics for the Mask R-CNN model and sampling technique

- Sampling Accuracy (Equation D-1) is a study-specific measure of the recognition quality of the sampling techniques calculated by finding the fraction of correctly predicted fully visible tubers out of all sampled tubers.

$$Accuracy = \frac{\text{Number of fully visible sampled tubers}}{\text{Total number of sampled tubers}} \quad \text{(D-1)}$$

- Recall is the ratio of all true positives to all predicted, which measures how well the machine learning model accurately predicts fully visible tubers. It is calculated using Equation D-2. In this research, true positive refers to the number of fully visible tubers that the model correctly predicted, while false negative is the number of tubers that were wrongly predicted to be partially visible.

$$Recall = \frac{\text{True positive}}{\text{True positive + False negative}} \quad \text{(D-2)}$$

- Precision is the ratio of all true positives (fully visible tubers predicted correctly) to all actual positives (all fully visible tubers present). It is calculated using Equation D-3. False positive is the number of tubers wrongly predicted to be fully visible.

$$Precision = \frac{\text{True positive}}{\text{True positive + False positive}} \quad \text{(D-3)}$$

- AUROC measures how well the model can discriminate between fully visible and partially visible tubers. The higher the AUROC, the better the sampling model is at predicting fully visible tubers as fully visible tubers and partially visible tubers as partially visible tubers. Mathematically, this can be computed with Equation D-

4, which estimates the area of the curve that plots the true positive rate (TPR) against the false positive rate (FPR) at various classification thresholds (t).

$$AUROC = \int_0^1 TPR\left(FPR^{-1}(t)\right)dt \quad \text{(D-4)}$$

- Average mIoU measures the similarity between the predicted segmentation masks and the corresponding ground truth masks. This metric assessed the Mask R-CNN model and the two sampling techniques. Equation D-5 shows how the IoU is calculated. The mIoU is the average of all detected tubers' IoUs in an image.

$$IoU = \frac{Area\ of Overlap}{Area\ of Union} \quad \text{(D-5)}$$

- Average precision evaluates the accuracy of the Mask R-CNN model in detecting potato tubers by measuring the detected tubers' precision and recall at different confidence threshold levels. The AP is computed as the area under the Precision-Recall (PR) curve, as seen in Equation D-6, constructed by plotting the precision values against the corresponding recall values at different IoU thresholds.

$$AP = \int_0^1 p(r)dr \quad \text{(D-6)}$$

where $p(r)$ is the precision at recall level r

Appendix E: Statistical properties of the image features

Table E-1: Summary statistics of the Hu moments features showing the threshold value for each feature.

Feature	Class	25th percentile	Median	75th percentile	Range	Mean	Point biserial correlation	Threshold value
H1	0	6.80 x 10⁻4	7.28 x 10^-4	8.04 x 10^-4	(0.63 - 1.46) x 10^-3	7.64 x10^4	-0.08	x
	1	6.96 x 10^-4	7.32 x 10^-4	7.73 x 10^-4	(0.64 - 1.00) x 10^-3	7.44 x 10^-4		
H2	0	0.49 x 10^-7	1.14 x 10^-7	2.25 x 10^-7	(0.05 - 121.00) x 10^-8	1.74 x 10^-7	-0.03	x
	1	0.88 x 10^-7	1.42 x 10 ^-7	2.05 x 10^-7	(0.91 – 61.08) x 10^-8	1.64 x 10^-7		
H3	0	0.43 x 10^-11	1.45 x 10^-11	4.09 x 10^-11	(0.09 - 717) x 10^-12	3.66 x 10^-11	-0.27	(0.13 – 0.61) x 10^-11
	1	0.13 x 10^-11	0.28 x 10^-11	0.61 x 10^11	(0.01 - 31.1) x 10^-12	0.47 x 10^-11		
H4	0	1.44 x 10^-13	8.56 x 10^-13	41.25 x 10^-13	(0.08 – 12238.55) x 10^-14	59.88 x 10^-13	-0.20	(0.76 – 4.84) x 10^-13
	1	0.76 x 10^-13	1.54 x 10^-13	4.84 x 10^-13	(0.11 – 627.44) x 10^-14	4.26 x 10^-13		

Feature	Class	25th percentile	Median	75th percentile	Range	Mean	Point biserial correlation	Threshold value
H5	0	-136.92 x 10^-28	935.60 x 10^-28	125.42 x 10^-25	(-735.32 – 15415.13) x 10^-24	240.73 x 10^-24	-0.09	x
	1	-9.90 x 10^-28	236.75 x 10^-28	2.89 x 10^-25	(-4.21 – 87.150 x 10^-24	1.23 x10^-24		
H6	0	-0.01 x 10^-15	0.01 x 10^-15	0.47 x 10^-15	(-245.85 – 607.12) x 10^-16	1.65 x 10^-15	-0.11	x
	1	-0.01 x 10^-15	0.01 x 10^-15	0.07 x 10^-15	(-3.93 – 31.34) x 10^-16	0.10 x 10^-15		
H7	0	-10.18 x 10^-25	0.01 x 10^-25	10.97 x 25^-25	(-276.42 – 1866.81) x 10^-23	16.21 x 10^-23	-0.05	x
	1	-0.70 x 10^-25	0.02 x 10^-25	0.62 x 25	(-0.98 – 1.40) x 10^-23	0.01 x 10^-23		

Table E-2: Summary statistics of the Edge and colour-based image features showing the threshold value for each feature.

Feature	Class	25th percentile	Median	75th percentile	Range	Mean	Point biserial correlation	Threshold value
S/N_1	0	0.60	0.67	0.73	0.36 -0.84	0.66	0.19	0.64 – 0.69
	1	0.64	0.69	0.74	0.48 -0.81	0.69		
S/N_2	0	0.73	0.76	0.80	0.50 -0.91	0.76	0.32	0.78 – 0.79
	1	0.78	0.79	0.81	0.73 -0.89	0.79		
Ellipse_ratio	0	2.32	2.69	3.19	1.45 - 3.92	3.05	-0.08	x
	1	2.08	2.34	2.64	1.41 - 3.51	2.38		
Ellipticalness	0	0.00	1.05	1.09	0.00 -1.29	0.76	0.28	1.03 - 1.05
	1	1.03	1.04	1.05	0.00 - 1.20	1.02		
Circularity	0	0.68	0.75	0.80	0.42 - 0.89	0.73	0.23	0.61 – 0.78
	1	0.74	0.78	0.82	0.61 - 0.88	0.78		
Convexity defect	0	558.50	1180.00	2251.25	238 - 8934.00	1673.87	-0.34	385.00 – 750.00
	1	383.50	545.50	747.00	269 - 2888.00	673.34		

Tables E-1 and E-2 summarise the fundamental statistical properties associated with the features alongside the corresponding threshold values chosen, where class 1 is fully visible, and class 0 is partially visible. The feature that demonstrated the most significant contribution in discriminating between fully visible and partially visible tubers was the convexity defect, exhibiting a correlation coefficient of -0.34. This negative correlation suggests that tubers with more convexity defects are more likely to be partially visible. A specific range was chosen to establish threshold values for each variable, typically between the 25th and 75th percentiles across the fully visible class. The aim was to identify values within this range that minimized the overlap between fully visible and partially visible tubers. For example, the threshold range for the "ellipticalness" feature was set between 1.03 and 1.05. This range captured 50% of fully visible tubers while encompassing 0% of partially visible ones.

Although the threshold range used for "ellipticalness" guaranteed 0% of partially visible tubers based on the validation dataset, depending solely on a single feature for tuber selection proved impractical, primarily because the correlation coefficient exhibited a maximum value of merely 0.34. More so, each feature had a unique contribution to the discrimination of the fully visible tubers from partially visible ones. For instance, the signal-to-noise ratios provide information about whether the tubers were fully visible or not and whether the tubers were also well segmented since poorly segmented tubers had considerably higher signal-to-noise ratios. As a result, a subset of seven features, selected based on their distinct discriminatory capabilities between fully and partially visible tubers, as indicated by their distributions and correlation coefficients, was utilized as

thresholding parameters. This approach aimed to improve the accuracy of the sampling process by effectively identifying and selecting well-segmented and fully visible tubers.

Appendix F: Evaluation metrics for the size estimation regression models.

- Root Mean Square Error (RMSE) measures the average deviation of the software-estimated tuber lengths from the measured values. It is calculated with Equation 4.4.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad \text{(F-1)}$$

where y_i is the observed value, $\hat{y}_i$ is the predicted value, and n is the number of observations.

- Normalized Root Mean Square Error (nRMSE) allows for comparing models in different scales. It helps to understand if the proposed method estimates the major diameter better than the minor diameter or vice versa. The formula for calculating the nRMSE is given in Equation F-2.

$$\text{nRMSE} = \frac{\text{RMSE}}{\bar{y}} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}i)^2}}{\frac{1}{n}\sum i=1^{n} y_i} \qquad \text{(F-2)}$$

where y is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the average of actual values, and n is the number of observations.

- The coefficient of determination (R2) is a statistical measurement that tells how the variation in one variable results from the variation in another. It is calculated as the ratio of the model square of squares (MSS) to the total sum of squares (TSS), as shown in Equation F-3.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}i)^2}{\sum 1i=1^{n}(y_i - \bar{y})^2} \qquad \text{(F-3)}$$

where SS_{res} is the sum of squared residuals, and SS_{tot} is the total sum of squares.

- Lin's Concordance correlation (CCC) is the product of the Pearson correlation coefficient between the measured and estimated lengths and a bias correction

factor that adjusts for the systematic difference between the measurements. It accounts for the variation and deviation of the measured and estimated lengths of the tubers, as shown in Equation 4.7.

$$\rho_c = \frac{2 \cdot \mathrm{cov}(y,\hat{y})}{\mathrm{Var}(y)+\mathrm{Var}(\hat{y})+(\bar{y}-\bar{\hat{y}})^2} \qquad \text{(F-4)}$$

where y is the vector of actual values, $\hat{y}$ is the vector of predicted values, $\mathrm{cov}(y, \hat{y})$ is the covariance between y and $\hat{y}$, $\mathrm{Var}(y)$ and $\mathrm{Var}(\hat{y})$ are the variances of y and $\hat{y}$, and $\bar{y}$ and $\bar{\hat{y}}$ are their means, respectively.