Modeling Physical Characteristics of Biological Systems using Quantitative Image Analysis and Machine Learning

By

ALEXANDER GEORGE OLENSKYJ
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOLOGICAL SYSTEMS ENGINEERING

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Gail M. Bornhorst, Co-Chair

J. Mason Earles, Co-Chair

Irwin R. Donis-González

Committee in Charge

2022

Acknowledgements

This dissertation and the studies described within would not have been possible without the support of my mentors, colleagues, friends, and family.

First, I would like to thank my advisors, Dr. Gail Bornhorst and Dr. Mason Earles for their guidance and support. Their insight and feedback were invaluable in helping me achieve both personal and professional growth. I would also like to thank Dr. Irwin Donis-González for his advice and contributions to the design and analysis of a large portion of this work as well as his service on my qualifying examination committee. Additionally, I would like to thank my collaborators Brent Sams and Pranav Raja for their advice and input.

I would like to acknowledge others who have contributed their time, expertise, and experimental support, including Tanya Garcia-Nolen, Dr. Yamile Mennah Govela, Dr. Zhenghao Fei, and Vishal Singh. I would also like to acknowledge E. & J. Gallo Winery for graciously providing the data and opportunity to carry out research in their fields. In addition, I would also like to thank Drs. Juliana De Moura Bell, Bruce German, Bryan Jenkins, and Shrinivasa Upadhyaya for serving on my qualifying examination committee as well as for sharing their knowledge and understanding with me throughout the past years, both inside and outside of the classroom. I would also like to acknowledge Sam Arcement for her support and her wisdom.

In addition to the research support, I would like to express my appreciation for the organizations who have helped fund this research, including the UC Davis Departments of Food Science, Biological and Agricultural Engineering, and Viticulture and Enology. I would also like to thank the external funding groups who provided support, namely the American Society for Enology and Viticulture as well as the USDA National Institute of Food and Agriculture.

Finally, I would like to thank all my colleagues and friends who I have met in my time carrying out these studies, including my lab mates and members of the BAE department who provided me with both essential guidance and encouragement as well as fond memories. Most importantly though, I would like to thank my parents, Paula Olenskyj and Andy Olenskyj, and my family for their constant love and encouragement. I would also like to thank Vivian Vuong for her valuable insight as well as her unwavering support and strength. I could not have achieved any of this without them and I am so grateful and fortunate to have had them by my side.

Abstract

Nondestructive imaging combined with recent advances in data processing techniques allows for minimally invasive, high-throughput, time-series analysis in both food and agriculture. The flexibility of neural network architectures allows for prediction of continuous output metrics, which may be valuable in analysis of both food and agriculture. In this work, nondestructive image analysis was used to demonstrate the relevance of image data to physical properties in food and agricultural systems as well as the potential for predictive models to increase the efficiency of quantitative data analysis.

Apples undergoing in vitro gastric digestion were used as a model food system, and X-ray micro-computed tomography data were shown to relate to peak force of apple tissue during compression. This relationship was probed further using a deep learning approach, where the compression curves of apple tissue during in vitro gastric digestion were predicted using a regression convolutional neural network (CNN) model. Results under cross-validation demonstrated strong accordance between predicted and measured compression curves, with an R^2 of 0.939 and RMSE of 4.36 N across measured force values. This relationship declined in samples from a holdout set, with an RMSE of 14.3 N, although this result was influenced strongly by the incubation medium tested (water vs. gastric juice).

Within the agricultural domain, the task of nondestructive yield estimation was demonstrated in vineyards. Images of grapevines collected using proximal sensing were associated with yield measured at harvest using a commercial yield monitor, allowing for a dataset of 23,581 yield measurements predicted using 164,699 images to be collected. Three deep learning architectures were used to predict the measured yield values from the images: object detection, CNN regression, and a transformer regression network. Regression-based architectures were used to eliminate the bottleneck of hand-labeling images. Results demonstrated that regression methods performed comparably to object detection methods without the need for hand labeling. Grape yield from within a training set, was correlated with model output at a mean absolute percent error of 6.4% when aggregated into 10 m regions using a transformer model. This study also demonstrated performance on a representative holdout set, with an error of 18% obtained using the same model and conditions.

Overall, this study demonstrated the potential of deep learning combined with nondestructive imaging for quantitative analysis of food and agricultural systems. Results demonstrated that image data can be related to mechanical properties of food materials as well as yield in an agricultural environment.

Contents

Figures

Tables

Equations

CHAPTER 1. INTRODUCTION

Throughout human history, mathematical techniques with theoretical applications have typically outpaced their accessibility. For example, the Radon Transform used to reconstruct computed tomography (CT) images was first described in 1917, but the first CT scanner was not developed until over half a century later, in 1972 (Kalender, 2006; Ramlau and Scherzer, 2018). Likewise, convolutional approaches to computer vision and pattern recognition were discussed as early as 1980, and almost 20 years later, in 1998, the technique was only viable for low-resolution 32x32 pixel images (Fukushima, 1980; LeCunn et al., 1998). Another 14 years passed before larger image sizes could be processed, demonstrated in 2012 by Krizhevsky et al (Krizhevsky et al., 2012). Although technological innovations will continue to emerge in the near and far future, recent advances in availability of computational resources and development of algorithms have made rapid and accurate analysis of biological systems formerly too complex to assess now possible (Kamilaris et al., 2017).

As an example, benchtop micro-CT instruments are now capable of resolving on the order of 10 µm, with scans taking place in less than an hour (Dhondt et al., 2010). The accessibility of these benchtop instruments coupled with the benefit of a nondestructive, quantitative 3D imaging platform provides opportunities for applications in biological sciences, where nondestructive data collection helps to reduce sample preparation time and variability in time-series data (Bultreys et al., 2016; Schoeman et al., 2016). Specifically in the field of food analysis, biological variability is a challenge, and procedures such as texture analysis require extensive subsampling within replicate measurements to get a reliable result (M. C. Bourne, 2002). This practice is both time and resource intensive, due to both the raw material and labor requirements.

Considering the efficiency of image data collection, the prospects of quantifying information from cameras represents a promising direction in areas where efficiency is particularly important. Specifically, the agricultural domain exemplifies an area with a high volume of raw material, small profit margins, and a large labor requirement. This labor requirement can make optimal crop management cost-prohibitive. For example, due to the labor-intensive aspect of vineyard management, typical operations are conducted on the "block" level at the smallest spatial scale, where a block can be as large as 30 acres (E.J. Gallo Winery, 2020). This gap between the level of spatial variability and the level of management practice represents an opportunity for growers to optimize their crop year-to-year (Priori et al., 2013). However, before management decisions can be made on such a fine spatial scale, growers need large-scale, high-resolution data. Camera sensors present an encouraging solution to this challenge. Cameras have become smaller and more prevalent, allowing for a data collection platform with the key properties of being high-throughput, minimally invasive, and inexpensive (Kwon and Park, 2017). As such, the throughput and cost advantages of consumer camera sensors as a data collection platform hold promise as tools for improving efficiency.

In both the utilization of time-series micro-CT in food analysis and implementation of a color imaging in agricultural environments, image analysis allows for collection of data which would otherwise be impractical or even impossible. While many food researchers have attempted to quantify changes during dynamic processes, such as digestion (Drechsler and Ferrua, 2015; Swackhamer et al., 2019), ripening (Volz et al., 2003), or storage (Eboibi and Uguru, 2017; Fortuny et al., 2003), these measurements have primarily used discrete samples for each time point and are manually performed. Likewise, agricultural researchers typically sample their crop during the growing season to assess ripeness and quality (Hellman, 2004; Zoui et al., 2010), but these discrete samples are time and labor-intensive to collect.

To mitigate those challenges, nondestructive imaging combined with recent advances in image processing techniques will allow for minimally invasive, high-throughput, time-series analysis in both food and agriculture. Some image-based interventions have been explored recently, such as in the nondestructive analysis of ice cream microstructure during storage (Pinzer et al., 2012), or the hyperspectral analysis of grapes on the vine (Aquino et al., 2018). However, recent advancements in the area of deep learning for image analysis have improved the robustness and accuracy of image analysis algorithms, and therefore hold promise in increasing the applicability for nondestructive image analysis (Lecun et al., 2015). Yet, a majority of image-based neural network tasks have revolved around image classification and object detection (Kamilaris and Prenafeta-Boldú, 2018). While these two tasks are useful and applicable to a wide array of fields, they are not typically quantitative or used to predict a continuous response variable, making them less suited in the measurement of food properties or the quantification of parameters important to the agricultural industry, such as evapotranspiration or crop yield. However, due to the flexibility of neural network architectures, prediction of continuous output metrics from image data inputs is possible, and the exploration of this use case may be valuable in the analysis of both food and agricultural systems (Häni and Roy, 2019).

In summary, new techniques in image data acquisition and interpretation have the potential to make large impacts in both the time-series food analysis and agricultural management spaces. In food analysis, collection of time-series data is time and resource-intensive due to the destructive nature of many measurement techniques and the high variability. Moreover, measurements such as hardness and moisture content determination tend to be lump-sum measurements, rather than spatially dependent metrics. In the agricultural domain, due in part to the technical and logistical challenges of data collection, current practices involve treating large regions of growing crops as a single unit, ignoring the differential needs of plants within that space. In both the food and agricultural domains, these limitations may be overcome through the use of quantitative imaging as a nondestructive, spatially resolved platform.

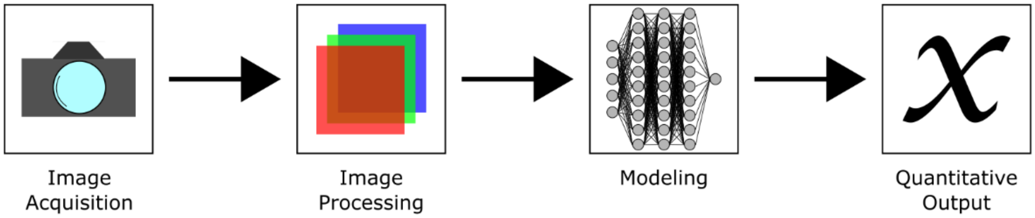


Figure 1.1. General pipeline for each study. Image acquisition involves collection of raw data. Image processing includes operations such as resizing, cropping, scaling, or otherwise modifying the images for input into the model. The model itself refers to the algorithm which accepts 2D or 3D images as an input and produces a quantitative output variable.

In the studies to follow, novel applications of recent techniques in the field of quantitative imaging and machine learning are discussed. Although the studies are conducted over a range of length scales, from micrometers to hectares, each study follows the same general pipeline (Fig. 1.1). Applications of micro-CT are presented first as an example of the advantages of image processing within a controlled experimental environment. Next, machine learning is assessed as a tool for analysis of CT data and quantification of continuous physical properties of food systems. Finally, to test the scope of applicability of image-based machine learning techniques in agricultural systems, novel image processing techniques applied to data collected with low-cost sensors in an uncontrolled agricultural environment is demonstrated. The overall objective of the dissertation was to demonstrate the applicability of nondestructive imaging for quantitative analysis to food and agriculture. In all cases, the rapid and nondestructive nature of the imaging technique was used to accelerate tasks which would formerly require extensive time or labor to perform. The specific objectives of this study are as follows:

1. Assess the effectiveness of nondestructive quantitative imaging for measuring physical properties in food systems during biological processing.
    a. Determine whether micro-CT imaging can be used to measure biophysical changes during biological processing, using in vitro gastric digestion of apple as a model system.
        - *Hypothesis: Micro-CT imaging can quantify changes in food structure over time, and the sensitivity will allow for significant differences between digestion treatments to be observed within the digestion time.*

    b. Compare changes observed using quantitative micro-CT imaging with destructive measurements of hardness and moisture content during in vitro gastric digestion.
        - *Hypothesis: Image-based metrics and destructive measurement of moisture content will demonstrate similar trends over the digestion time.*
2. Evaluate the synergy between quantitative imaging and machine learning for nondestructive prediction of mechanical properties of food systems.
    a. Leverage micro-CT imaging for calculation of spatially resolved food structure by way of compression curves without destruction of samples, using in vitro gastric digestion of apple as a model system.
        - *Hypothesis: Compression profiles of apple tissue can be estimated directly from 3D CT data using an end-to-end machine learning approach.*

    b. Explore the feasibility of machine learning as a high-throughput method for analyzing time-dependent trends in food processing.

- *Hypothesis: Robust machine learning algorithms trained on 3d CT data will be capable of depicting time-dependent softening of apples during in vitro gastric digestion using.*

3. Investigate the effectiveness of nondestructive quantitative imaging in combination with machine learning for acceleration of human tasks in viticulture.

   a. Develop a low-cost data collection and analysis pipeline using off-the-shelf hardware to generate geospatially resolved image data in vineyards and use image data to predict grape yield nondestructively.

      - *Hypothesis: Deep learning models trained on images of grapevine canopies will be able to use image features to predict grapevine yield, either by localization of grapes within the canopy or by directly predicting yield from an input image.*

   b. Explore the practicability of using end-to-end models designed to predict yield directly from image data, as opposed to the more conventional object detection.

      - *Hypothesis: End-to-end models will allow for higher quality yield predictions without the need for hand-labeling or manual model tuning.*

CHAPTER 2. LITERATURE REVIEW

Computational power and resource availability is beginning to outpace the complexity of algorithms developed for data analysis, and these advancements in accessible computing resources have led to rapid development of new techniques and applications (Li and Chen, 2014). In the application of new computational methods, one of the major developments has been the design of tools for processing large amounts of data. One of the most impactful developments has been techniques which leverage graphics processing units (GPUs) for massively parallel computation (Raina et al., 2009). As a result, in the field of computer vision, datasets on the order of millions of images are widely accessible (Krizhevsky, 2009; Lin et al., 2014; Russakovsky et al., 2015), and leveraging data on this scale is now possible. The accessibility of large amounts of data and the models to understand it are driving innovation in domains like food and agriculture, where emerging techniques such as computer vision (CV) and machine learning (ML) are being applied to numerous areas of interest.

Specifically, there has recently been increased attention on the functional aspects of foods: health benefits such as satiety, sustained nutrient release, or enteric delivery of acid-unstable probiotics or nutrients (Dye and Blundell, 2002; Halford and Harrold, 2012; Mattila-Sandholm et al., 2002; Norton et al., 2014). The food functional properties can be diverse; thus, thorough characterization of food before consumption and during digestion is essential for the understanding of potential biological function of the food under study. While there are currently methods which allow for characterization of food functionality, methods for food characterization and analyses of the subsequent data can be made more resource and time efficient with CV and ML methods.

In addition to the potential for application of CV and ML methods in food analysis, the agricultural field also represents an area which may benefit from resource and time efficiency improvements possible with new computational resources. Historically, the agricultural sector has been fast to incorporate new technological advancements. Although conventional methods of food production had sustained humanity for thousands of years, the rising population in the middle of the 19th century led to the rapid adoption of technology by the food and agricultural sector, which has allowed crop yield per unit area to grow each year since (Borlaug, 2002). Now, as the global population continues to rise exponentially, the pace of development in agricultural technology has continued to accelerate (W. J. Chancellor, 1981), leading most recently to a widespread interest and adoption of CV and ML techniques (Kakani et al., 2020; Kamilaris and Prenafeta-Boldú, 2018). Specifically, one notable area of interest for growers is yield prediction, which offers considerable benefit to growers. For example, in viticulture, information about expected crop yield allows growers to adjust crop management via measures like cluster thinning to improve grape quality (De La Fuente et al., 2015). Additionally, accurate yield estimation allows for labor and storage considerations to be considered early, to avoid excess waste and lost profit (Nuske et al., 2014b).

In this review, recent advances in computer vision and machine learning will be outlined. Additionally, the state of the art for quantitative analysis in food structural breakdown and viticultural yield estimation will be examined, along with areas where CV and ML methods may provide benefit.

2.1. Recent Advances in Computer Vision and Machine Learning

Since the inception of CV and machine learning ML, which occurred at roughly the same time (Frank Rosenblatt, 1958; Roberts, 1963), the power and application of these combined methods has greatly expanded. Until recently, ML has been roughly divided into two complementary tasks. *Feature extraction*, where input information is processed to remove irrelevant data and produce explanatory numerical representations, can be followed by *pattern classification*, where features are used to produce a desired response, such as assignment of a class value to the input (Duda et al., 2000; Nixon and Aguado, 2002). While these steps are commonly optimized together, a theoretically perfect feature representation of an input would be simple to classify; likewise, a perfect classification algorithm would work even with low quality features (Duda et al., 2000). Considering these principles, a subset of ML known as deep learning (DL) follows the first notion in which feature extraction is the primary focus using a technique called representation learning. In DL, the process for representation learning is driven by the backpropagation method (LeCun et al., 1989), where a gradient of an objective function is calculated with respect to the parameters within the model used to produce the result. Backpropagation allows for optimization by adjustment of the parameters of an arbitrarily large model to generate ideal representations of input data via simple differentiation and iteration, as long as the model is differentiable with respect to an objective function (Lecun et al., 2015).

While the concept of representation learning via backpropagation has existed for decades, only recently have these methods seen widespread use, primarily due to the design of techniques leveraging GPUs for efficient image processing and computation of gradients (Krizhevsky et al., 2012; Raina et al., 2009). Primarily, the method which has been employed most often for efficient computation on GPUs is the convolutional neural network, or CNN, which makes use of learned kernels of sizes ranging from 3x3 to 7x7 pixels applied to images via a sliding window approach (Szegedy et al., 2015). These kernels allow for regional information to be incorporated into increasing levels of abstraction. Additionally, processing of large images can be done efficiently, without requiring a large number of model weights, since the same small kernel is applied to the entire image (Lecun et al., 2015; LeCun et al., 1989). The convolutional approach led to an explosion in vision-based models in the last 10 years (Dhillon and Verma, 2020) (Table 2.1). Recently however, a new model architecture known as the transformer has emerged as a powerful model for vision-based tasks. The transformer model does not explicitly make use of convolutions, but instead uses attention mechanisms to incorporate regional-based information, requiring a larger amount of memory to do so as compared with a typical CNN (Carion et al., 2020; Dosovitskiy et al., 2020; Vaswani et al., 2017). Leveraging both convolutional and attention-based techniques, vision-based deep learning methods have been developed for a variety of tasks, including image classification, object detection, image segmentation, instance masking, image generation, and more (Table 2.1). Of these tasks, image classification and object detection are some of the most common, due to their wide area of application (Guo et al., 2016).

Image classification models are trained to output a vector representing the probability of an input image belonging to a set of possible classes. During training, models are generally trained by minimizing the negative log likelihood of predicting the correct class (Shrestha and Mahmood, 2019). However, the negative log likelihood objective function can be replaced by a different objective function, such as mean squared error (MSE) for regression tasks (Fu et al., 2018). Additionally, removal of the final classification layer entirely leaves only the feature extraction stage of the model. This allows the image classification architecture to be used in more complex models, such as the Single Shot Detector (SSD) model, which leverages the Visual Geometry Group (VGG) backbone for object detection (Table 2.1) (W. Liu et al., 2016).

SSD and other object detection models are therefore models with increased complexity as compared with image classification models. Most object detection models have two stages: one stage for proposing regions where objects may be present, and another stage for classifying and more accurately localizing objects in the proposed regions (Wu et al., 2020). Classification is performed the same way as in image classification models, where a vector of class probabilities is used along with a negative log likelihood objective function. For localization, object detection models are trained using a regression loss function such as MSE on a vector of four values representing the location of the bounding box (x and y coordinates along with box dimensions, for example) (Ren et al., 2017). The additional complexity created by these two stages and multiple objective functions generally leads to models which are slower to both train and use for prediction. However, some approaches, such as those used by the You Only Look Once (YOLO) and SSD models, skip the region proposal step and simply use the same set of regions for any input image (Table 2.1). This leads to models with increased speed, at the expense of a small amount of accuracy (W. Liu et al., 2016; Redmon and Farhadi, 2018).

Other common models include those trained for the task of segmentation. These models are typically trained to output a pixel mask with a height and width equal to the input image size, with each (x, y) pixel location represented by a vector of class probabilities, similar to a classification task, but at a pixel level (Minaee et al., 2021). Segmentation models are generally trained using an objective function which rewards each pixel for which the correct class is predicted and penalizes pixels where the model does not select the correct class, such as the Generalized Dice Loss (Sudre et al., 2017). Instance segmentation represents a particularly challenging subset of both general segmentation as well as object detection models, where models are trained to detect and classify objects in an image, and additionally output pixel-level masks for each distinct object (Hafiz and Bhat, 2020). These models, such as Mask R-CNN, use multiple loss functions to represent the detection, classification, and masking components which comprise the task of instance segmentation (He et al., 2017).

Finally, the creation of the generative adversarial network (GAN) has opened opportunities for the use of deep learning in the creation of images, rather than analysis alone (Goodfellow et al., 2014). GAN models are trained using two separate networks: a generator and a discriminator network. The generator is trained to generate an image similar to images within a dataset and the discriminator is trained to operate on a set of both real images from the dataset along with generated images, with the goal of detecting the real and fake examples. The generator is penalized for producing images which the discriminator can easily pick out. Conversely, the discriminator is penalized for failing to spot the fake images. After training, the discriminator is typically not required, and the generator is used on its own to produce realistic but fake images which are difficult to differentiate from real-life examples. There are many examples of GAN models (Fei et al., 2021; Isola et al., 2017; Karras et al., 2020), but the primary novelty lies in the adversarial training procedure, as opposed to the specific architectural implementation.

Table 2.1. Notable deep learning architectures for vision-based tasks. Model performance on image classification tasks reflects ImageNet top-1 performance in percent (Deng et al., 2009). Performance for object detection and instance segmentation models is given as box AP on COCO (Lin et al., 2014). Segmentation performance listed represents Cityscapes mean IoU in percent (Cordts et al., 2016). Gaps in the table represent unreported values.

Task	Architecture	Novelty Introduced	Parameters (Millions)	Performance	References
Image Classification	AlexNet	GPUs for training models on huge datasets	60	63.3	(Krizhevsky et al., 2012)
	VGG (Visual Geometry Group)	Simple but very deep model with small (3x3 convolution) for high performance	144	74.5	(Simonyan and Zisserman, 2015)
	GoogLeNet	Multi-scale convolutional kernels learned simultaneously	5	66	(Szegedy et al., 2015)
	ResNet (Residual Network)	Residual connections for training deeper networks	60	78.6	(He et al., 2016)
	MobileNet	Depthwise separable convolutions for parameter reduction	5.4	75.2	(Howard et al., 2019, 2017; Sandler et al., 2018)
	EfficientNet	Neural architecture search for network optimization via width and depth scaling	120	87.3	(Tan and Le, 2019, 2021)
	ViT (Vision Transformer)	Attention mechanisms in image classification	304	85.3	(Dosovitskiy et al., 2020)
Object Detection	R-CNN (Region-based CNN)	Region proposal followed by classification using CNNs	53	34.9	(Girshick, 2015; Girshick et al., 2014; Ren et al., 2017)
	YOLO (You Only Look Once)	Object detection at multiple scales with residual connections and upsampling steps but no learned region proposals	65	33	(Redmon et al., 2016; Redmon and Farhadi, 2018, 2017)
	SSD (Single Shot Detector)	Object detection at multiple scales of the VGG backbone with no learned region proposals	26.3	24.4	(W. Liu et al., 2016)
	DETR (Detection Transformer)	Attention mechanisms in object detection	60	44.9	(Carion et al., 2020)
Segmentation	U-Net	Encoder-decoder for learned multi-step downscaling and upscaling with shared information (skip connections)	7.7	--	(Weng and Zhu, 2015)
	FCN	Encoder for information extraction followed by	134	65.3	(Long et al., 2015)

	(Fully Convolutional Network)	upscaling for pixelwise segmentation			
	DeepLab	Deep encoder-decoder network with either conditional random fields applied to upscaled feature maps (V1 and V2) or atrous convolution for multi-scale contextualization (V3)	60	81.3	(Chen et al., 2017, 2018, 2015)
Instance Segmentation	Mask-RCNN	Pixelwise segmentation within learned region proposals	55	37.1	(He et al., 2017)
	PA-Net (Path Aggregation Network)	Improvement on Mask-RCNN via optimization of information flow by combining abstract and low-level features within the model	--	42	(Liu et al., 2018)
	YOLACT (You Only Look at Coefficients)	Real-time instance segmentation via generation of whole-image masks and instance weights applied to generated masks	--	34.1	(Bolya et al., 2020, 2019)
Image Generation	GAN (Generative Adversarial Network)	A class of architectures based around iterative optimization of one network to generate data and another to differentiate real from generated data to learn to generate increasingly realistic data	--	--	(Goodfellow et al., 2014)

2.2. Analysis of Food Structure and its Breakdown in Digestion: Conventional Methods and Areas for Application of CV and ML

One unique aspect to the study of food structure is that unlike most other physical objects in our daily lives, food is created with a structure which is meant to be destroyed. In this case, the destruction is carried out by the human digestive system, and the destruction is necessary for delivery of nutrients to the body. However, the rate, location, and time of delivery of nutrients is a key way in which foods can exhibit functionality (Norton et al., 2014). This functionality is driven by the interaction of food structure and composition with the human digestive system.

The human digestive process can be broken into four major stages: mastication, gastric digestion, small intestinal digestion, and large intestinal fermentation (Bornhorst and Singh, 2014). Mastication and gastric digestion serve to break down food materials for efficient nutrient uptake in the small intestine (Parada and Aguilera, 2007). Indigestible and undigested material which leaves the small intestine is then fermented by the resident microbiota and the remaining material is excreted as waste (Singh et al., 2015).

Although most nutrient uptake occurs in the small intestine, efficient physical and chemical breakdown of food microstructure in the mouth and stomach is essential for releasing nutrients from the food matrix (Singh et al., 2015). In this context, food microstructure will be defined as food matrix components which exist on the order of 0.1 to 100 µm (Aguilera, 2005). Some examples include porous matrices such as those found in fruits, vegetables, and processed grain products, cell packing in natural foods, extracellular tissue structure, and emulsion composition. Without efficient microstructural breakdown, research has shown that nutrients will not leave the food matrix and cannot be taken up by the body (Grundy et al., 2016a; Parada and Aguilera, 2007). For this reason, if nutrient delivery is a desired outcome, an understanding of food microstructure and its interaction with the mastication and gastric digestion processes is essential for both the design of functional foods and the identification of potential functional properties of natural foods (Capuano and Janssen, 2021; Li et al., 2021; Norton et al., 2014). As an example, studies showing that processing of almonds is highly related to the release of lipid from the cellular matrix has led some researchers to question the validity of nutrition facts labels for almonds (Grundy et al., 2016b; Novotny et al., 2012). Furthermore, studies focusing on the extent of microstructural breakdown of almonds have been used to support claims that due to their minimal breakdown, almonds provide fewer calories than their current label suggests (Grassby et al., 2014; Grundy et al., 2016b, 2015).

2.2.1. Analysis of Food Structural Breakdown

This understanding of food microstructure and its breakdown during digestion typically requires qualitative and quantitative characterization of food systems, which can be achieved with a wide variety of measurement techniques (Nielsen, 2010). Qualitative characterization has included observations such as visual comparison of cell wall integrity before and after digestion, (Chen et al., 2011; Mennah-Govela and Bornhorst, 2016a) as well the appearance of contents in food particles as judged via localization of selectively stained material before and after digestion (Chen et al., 2011; Grundy et al., 2016a). Quantitative characterization performed in digestion research has included measurement of physical resistance to stress (Drechsler and Ferrua, 2015; Kong and Singh, 2009a; Nadia et al., 2021; Olenskyj et al., 2020; Somaratne et al., 2020), determination of moisture uptake (Kong and Singh, 2009a; Mennah-Govela and Bornhorst, 2016b; Nadia et al., 2021; Swackhamer et al., 2019), and measurement of specific nutrients before and after the digestion process (Liang et al., 2012; Luo et al., 2015; Opazo-Navarrete et al., 2018; Qiu et al., 2012).

Although these methods are widely used, challenges still arise in the study of temporal processes such as digestion. For qualitative measurement of food microstructure, optical techniques like visible microscopy or confocal laser scanning microscopy (CLSM) have been used previously to assess cellular disruption in materials like sweet potatoes, peanuts, protein gels, and almonds (Chen et al., 2011; Kong and Singh, 2009a; Mennah-Govela and Bornhorst, 2016b, 2016a; Opazo-Navarrete et al., 2018). In these studies, other than providing a visualization of the appearance of food, the microscopy offered little to no quantifiable results. Moreover, the procedural requirements for these imaging procedures can be extensive. In Chen et al. (2011), although no staining was required, the authors embedded particles in gelatin for imaging, which required time to set. In Mennah-Govela and Bornhorst (2016a), samples were first fixed in formalin for four days, then dried in ethanol and embedded in paraffin wax. Samples were then cut with a microtome and left in a slide incubator for up to an additional day. In addition to representing time consuming procedures, extensive preparation can alter the structure of the material, as demonstrated by Rodgers et al. (2022, 2021), in which researchers conducted similar steps to prepare a mouse brain for histology: formalin fixation, ethanol drying, and paraffin embedding. Between each step, the authors leveraged synchrotron micro-computed tomographic imaging, during which the 3D structure of the brain was captured in less than 30 minutes. X-ray images demonstrated that quantifiable tissue changes took place during sample preparation, and the changes were not isometric throughout the material (Rodgers et al., 2022).

Alternatively, as opposed to focusing on microstructure, some works focus instead on bulk mechanical properties to draw conclusions about microstructural behavior (Drechsler and Bornhorst, 2018; Liu et al., 2021). In these works, changing food texture during digestion (softening) or rheological properties were assessed to gauge the rate of the digestion process. However, textural metrics like maximum force at compression are scalar values measured on bulk solids which can be difficult to interpret with respect to mechanisms on the micro scale. For example, Liu et al. (2021) used texture analysis and rheometry to analyze the hardness, gumminess, shear, and frequency data of protein gels, along with SDS-PAGE analysis to study protein hydrolysis. Using this information, the authors proposed a mechanism of gel formation and breakdown during digestion. However, the detailed microstructure was unable to be observed directly, as the mechanical and chemical analyses performed were destructive and did not allow for subsequent visualization of microstructure (Liu et al., 2021).

*Table 2.2. Representative papers on analysis of microstructure of solid food during digestion and other time-series processes. Rows with a * represent studies which appear under multiple categories, where multiple properties of the food material were assessed.*

Category	Process	Material Assessed	Instrument Used	Analysis Output	Reference
Microscopy	Gastric Digestion	Almond	Scanning electron microscopy	Qualitative	(Swackhamer et al., 2019)
		Almond	Light microscopy, Transmission electron microscopy	Qualitative*	(Kong and Singh, 2009a)
		Peanut	CLSM	Qualitative	(Chen et al., 2011)
		Sweet potato	Light microscopy	Qualitative	(Mennah-Govela and Bornhorst, 2016a)
		Sweet potato	Light microscopy	Qualitative	(Mennah-Govela and Bornhorst, 2016b)
	Gastrointestinal Digestion	Almond	Light microscopy	Qualitative	(Grundy et al., 2016a)
	Mastication	Almond	CLSM	Qualitative	(Grundy et al., 2015)
Mechanical Testing	Aging	Cheese	Texture analyzer	Hardness*	(Vásquez et al., 2018)
	Cooking	Sweet potato	Texture analyzer	Hardness	(Mennah-Govela and Bornhorst, 2016a)
	Gastric Digestion	Almond	Texture analyzer	Compression curve*	(Kong and Singh, 2009a)
		Apple	Texture analyzer	Hardness*	(Olenskyj et al., 2020)
		Carbohydrate-based foods	Texture analyzer	Hardness	(Drechsler and Bornhorst, 2018)
		Potato	Texture analyzer	Failure strength, Toughness, Apparent Elasticity, Hardness	(Drechsler and Ferrua, 2015)

		Protein gel	Rheometer	Flow behavior, storage and loss moduli	(Liu et al., 2021)
		Protein gel	Texture analyzer	Hardness	(Somaratne et al., 2020)
		Wheat and rice products	Texture analyzer. Rheometer	Hardness, Flow behavior, Storage and loss moduli	(Nadia et al., 2021)
	Gelation	Protein gels	Texture analyzer	Hardness	(Opazo-Navarrete et al., 2018)
Nondestructive Imaging	Aging	Cheese	Hyperspectral camera	Spectral reflectance*	(Vásquez et al., 2018)
	Bubble growth	Dough	Synchrotron X-ray	Image intensity, Porosity	(Turbin-Orger et al., 2015)
		Dough	X-Ray micro-CT	Image intensity, Porosity	(Trinh et al., 2013)
	Drying	Tarkhineh	Color camera	Geometry, Image texture	(Ghaitaranpour et al., 2017)
	Foam decay	Milk	Synchrotron X-Ray	Image intensity, Porosity	(Eggert et al., 2014)
	Frying	Chicken nugget	Hyperspectral camera	Image texture	(Qiao et al., 2007)
	Frying, storage	Potato, Chocolate	Color camera, light microscopy	Image texture	(Quevedo et al., 2002)
	Gastric Digestion	Apple	X-Ray micro-CT	Image intensity, Porosity*	(Olenskyj et al., 2020)
	Staling	Bread	Color camera	Image texture	(Nouri et al., 2018)

For more holistic analysis of breakdown, microscopy can be coupled with quantitative metrics obtained via mechanical testing (Beaulieu et al., 2001; Kong and Singh, 2009a; Mennah-Govela and Bornhorst, 2016b), particle size analysis (Chen et al., 2011), moisture and pH measurement (Mennah-Govela and Bornhorst, 2016a), or chromatographic analysis (Opazo-Navarrete et al., 2018). The interaction between macroscale mechanical properties and the microscale can be important, such as when disruption of food microstructure on the order of micrometers in the form of moisture uptake influences a food particle’s tendency to break apart on the macro scale (Kong and Singh, 2009a). Conversely, in certain food systems, fracture that occurs on the macro scale increases surface area and therefore allows for an increased extent of reactions that occur on the micro scale, such as enzymatic or hydrolysis reactions, which may lead to cell rupture (Bornhorst et al., 2016). However, although qualitative and quantitative measurements can be coupled, the conventional analyses for these properties are destructive, meaning multiple observations cannot be made on the same sample.

In summary, within the study of food digestion, no single measurement technique can effectively characterize the entire concert of processes occurring as food is digested. Still, in experiments which couple qualitative and quantitative metrics for a more holistic analysis of digestion, due to the destructive nature of each of these measurements, time-series analysis of a single sample is impossible.

2.2.2. Role of CV and ML in food analysis

To solve the problem wherein time series analysis of a single sample is prevented due to the destructive nature of conventional methods, the combination of CV and ML has the potential to bridge the gap between qualitative and quantitative measurements, allowing for more powerful and efficient analysis. For example, one potential area for CV and ML to aid in research would be to quantify or classify food textural information from image data, allowing for both qualitative visual observations to be coupled with quantitative assessment. Variation in patterns of intensity and color within an image, known as image texture, can be a rich source of information about the sample under study (Nixon and Aguado, 2002). Specifically, quantification of image texture has been used previously to nondestructively quantify changes in food materials during processes such as frying and chocolate bloom (Quevedo et al., 2002), bread staling (Nouri et al., 2018), bread drying (Ghaitaranpour et al., 2017), chicken frying (Qiao et al., 2007), and cheese ripening (Vásquez et al., 2018). In these studies, nondestructive imaging allowed the researchers to follow a single sample during a dynamic process and make continuous quantitative measurements, allowing for quality assessment without excess waste (Table 2.2). Similarly, application of a ML classifier to image features has enabled some researchers to perform tasks on food samples such as quality assessment and even food analysis (Table 2.3). In particular, Qiao et al. (2007) used leave-one-out cross validation to evaluate the predictive capacity of a model which implemented a multilayer perceptron classifier for prediction of mechanical properties of chicken nuggets from color imagery. The authors were able to relate predicted and measured properties with an R^2 of between 0.62 and 0.7.

Table 2.3. Representative studies on computer vision and machine learning in food and postharvest technology. Features Extracted and Classification Method columns are combined for deep learning studies, as the models are used for both purposes.

Features Extracted	Classification Method	Task	Material Assessed	Instrument Used	Predicted Value	Reference
Grey level co-occurrence matrix (GLCM) features	Linear correlation	Food analysis	Bread	Color camera	Moisture, Firmness, Springiness, Consumer rejection	(Nouri et al., 2018)
GLCM features	Multilayer perceptron (MLP)	Food analysis	Chicken nugget	Color camera	Hardness, Toughness, Energy to break point	(Qiao et al., 2007)
GLCM features	Linear correlation	Food analysis	Tarkhineh	Color camera	Moisture Content	(Ghaitaranpour et al., 2017)
Fast Fourier Transform (FFT), Histogram analysis	K-nearest neighbor, MLP	Defect detection	Apple	X-Ray	Level of defect	(Kim and Schatzki, 2000)
Local binary patterns (LBP), Gabor filters, FFT, Texture,	Linear discriminant analysis (LDA),	Quality assessment	Carrot	X-Ray CT	Undesirable fibrous tissue class	(Donis-González et al., 2016)

Contrast, Intensity Features	Quadratic discriminant analysis (QDA), Mahalanobis distance, MLP					
LBP, Gabor filters, Texture, Contrast, Intensity Features	LDA, QDA, Mahalanobis distance, MLP	Quality assessment	Chestnut	X-Ray CT	Quality class	(Donis-González et al., 2014)
Geometric, Elliptical, Fourier descriptors, Invariant geometric moments, Color, Statistical textures, Filter banks, Invariant color moments	Multi-class support vector machine with radial basis function	Quality assessment	Corn tortilla	Color camera	Hedonic sub-class	(Mery et al., 2010)
Deep learning (VGG)		Classification, Food analysis	Food photos	Color camera	Food category, calorie count	(Ege and Yanai, 2017)
Deep Learning (Faster R-CNN)		Classification, Food analysis	Food photos	Color camera	Food category, calorie count	(Liang and Li, 2017)
Deep learning (Customized GoogLeNet)		Classification	Food photos	Color camera	Food category	(C. Liu et al., 2016)
Deep learning (ResNet)		Classification	Food photos	Color camera	Food category	(Kaur et al., 2019)
Deep learning (Mask R-CNN)		Defect detection	Apple	Color camera	Decayed tissue localization	(Stasenko et al., 2021)
Deep learning (YOLO, SSD)		Defect detection	Apple	Color camera	Defect category of image object	(Valdez, 2020)
Deep learning (VGG)		Classification	Date	Color camera	Ripening stage	(Nasiri et al., 2019)
Deep learning (Mask R-CNN)		Defect detection	Strawberry	Color camera	Bruised tissue localization	(Zhou et al., 2021)

In addition to experimental measurements, CV and ML are extensively used in the post-harvest industry, where combining image textural features with classification-based machine learning has been implemented for categorizing foods based on quality attributes such as fluid buildup (Kim and Schatzki, 2000), damaged tissue (Donis-González et al., 2013), and discoloration (Mery et al., 2010). In particular, Kim and Schatzki made use of X-ray imaging for classification of watercore in *Red Delicious* apples into three classes to obtain a performance of over 60% (Kim and Schatzki, 2000). The authors utilized 2D medical X-ray imaging as a nondestructive method of classifying food samples that have varying properties.

While these existing studies have demonstrated success using more conventional CV and ML methods, the representation learning approach offered by deep learning may allow for even more complex inferences to be made from image data (Table 2.3). Specifically, several food classification datasets composed of between 4,350 and 256,000 images categorized into between 6 and 520 individual food classes have emerged in recent years (Kaur et al., 2019). These datasets permit the development of models which can be used for more efficient food journaling for dietary and medical purposes via classification of food based on an image. More than just classification, datasets which include images of food along with their caloric value have also been developed for similar purposes (Ege and Yanai, 2017; Liang and Li, 2017). Caloric prediction represents an example of a regression task (prediction of a continuous as opposed to a discrete value), although the flexibility of DL architectures provides multiple avenues for approaching the problem. For example, Liang and Li (2017) used an object detection model (Faster R-CNN) to locate the food and a phantom placed in the image (a coin) to estimate food volume. Calorie count was then estimated based on caloric content of the predicted food per unit volume multiplied by the predicted volume. On the contrary, Ege and Yani (2017) took a different approach and trained a model end-to-end to output the caloric content of an image of food along with a predicted food class simultaneously. The model was then optimized based on an objective function which combined performance on both classification and regression tasks.

In summary, while the study of food digestion and other time-series processes is important, the conventional methods used for food analysis tend to prevent repeated observations made on a single sample, leading to inefficient use of time and resources. However, there are numerous applications of CV and ML within the food and postharvest industries which extract relevant information from image data in a nondestructive fashion. More recently, the rise of DL has allowed for the creation of even more accurate predictive models with a broad range of applicability within the food domain. Considering the power and flexibility of techniques like ML and DL, there is an opportunity for these approaches to be applied towards overcoming the challenges of time-series analysis in the food industry, but future research is warranted.

2.3. Yield Estimation: Conventional Methods and Areas for Application of CV and ML

While CV and ML have extensive applicability and promise in the food and postharvest industries, their use in an agricultural context, before food is harvested, is also well-established. In particular, yield estimation is a use case which has seen extensive development in recent years and may also benefit from recent advances in DL. The interest in this task has been driven by inefficiency in current industry practice. For example, one of the most widely used methods of viticultural yield estimation involves manually sampling from a small percentage of crops (e.g., approximately 1%) and extrapolating the distribution of yield data to the entire field. However, although this method is popular due to its low cost and complexity, manually sampling from vines requires a large labor investment, and the results can still be imprecise (Liu et al., 2020). The American Society for Enology and Viticulture produced a set of general recommendations regarding manual sampling in 1992 (Wolpert and Vilas, 1992) which include first counting the clusters in 10 or more vines spaced randomly throughout the field to estimate cluster count per vine. The second step involves individually weighing 10 or more randomly acquired grape clusters, followed by collection of over 200 additional clusters for weighing together. The authors concede that many vineyards will have to make concessions regarding the sample quantity, as the cost of the procedure may not be within a grower's budget (Wolpert and Vilas, 1992). More recently, one study on counting grape bunches demonstrated that measuring 15 vines took 1.5 hours and led to a 5% error in measuring yield on the same 15 vines, without attempting to extrapolate to other vines (Wulfsohn et al., 2012). Finally, as an additional point of reference, in the grape industry, yield prediction error has been reported to be accurate to within 30% using manual sampling (Sun et al., 2017).

Due to the labor requirement and inaccuracies of the manual sampling methods, techniques such as remote and proximal sensing have been applied to yield estimation to help increase the efficiency and accuracy of the process. Remote sensing can be conducted in two primary ways: via satellite or via unmanned aerial vehicles (UAVs). While satellite sensing can be conducted rapidly over a large area due to their large field of view, the tradeoff in satellite remote sensing is in the resolution of the data, which is low, typically between 10 and 30 m^2 (Khaliq et al., 2019). As a result, gaps between crop rows are typically averaged into the generated pixels of the resulting map. Cloud cover can also influence results, as satellites orbit at high altitude where clouds can occlude the ground. As opposed to satellite-based methods, remote sensing using UAVs can provide increased resolution, but the method requires a trained pilot to conduct the imaging procedure (Khaliq et al., 2019; Yang et al., 2019). In both satellite and UAV methods, for specialty crops with foliage such as tomatoes and grapes, the overhead angle and low resolution of remote sensing images typically prevents direct visualization of the harvestable portion of the crop in the image data (Di Gennaro et al., 2019). Instead, vegetation indices such as normalized difference vegetative index (NDVI) and leaf area index (LAI) are used for association with relevant agricultural parameters such as yield, grapevine vigor, or fruit quality (Anastasiou et al., 2018; Kazmierski et al., 2011; Sun et al., 2017).

Remote sensing in this case is an example of nondestructive imaging, and the vegetative indices extracted from image data represent features extracted for modeling. Specifically, indices are plotted against known values of yield or other quality parameters to produce a calibration which can be used for estimating yield based on indices of future unseen samples. These correlative approaches have been shown to perform well in association of remote sensing imagery with quality parameters such as total soluble solids, pH, and mechanical properties of berries (Anastasiou et al., 2018). Additionally, in other grain crops, such as corn and soy (Johnson, 2014) as well as cereals, wheat and barley (Panek and Gozdowski, 2021) remote sensing-based measurement of vegetative indices such as NDVI and land surface temperature and subsequent correlative modeling has been used extensively for yield estimation, with correlations demonstrating an R^2 value of between 0.61 and 0.77 at a resolution ranging from 250 m (Johnson, 2014) to the entire country scale (Panek and Gozdowski, 2021). However, in grapevines, properties of the foliage do not always correlate well with vine yield in terms of mass, and remotely-sensed yield estimates are prone to error (Anastasiou et al., 2018; Sun et al., 2017). Therefore, remote sensing studies aiming to predict yield in vineyards have had only limited success, with correlative R^2 values from satellite-based studies ranging from as low as effectively 0 (Anastasiou et al., 2018) to as high as 0.59 at a 30 m resolution (Sun et al., 2017). UAV studies have demonstrated an increased correlation, but results were variable, with studies demonstrating correlation fits with an R^2 of between 0 and 0.95 (Table 2.4). Notably, the best fit model employed a multilayer perceptron model along with vegetation indices to increase performance (Ballesteros et al., 2020). Still, in most cases, the overall inconsistency in the performance of UAV and satellite studies likely arises from error in the relationship between vegetative indices and yield. Proximal imaging, on the other hand, seeks to account for this error by imaging the fruit directly.

Table 2.4. Recent works on proximal, UAV, and satellite imaging for yield estimation with associated experimental conditions. Performance metrics listed are for prediction of mass per vine in proximal studies and either mass per vine or yield density (mass per unit area) in UAV and proximal studies. Ground resolution of UAV and satellite imaging studies are included under the Acquisition Method column. Notably, one UAV study has a resolution of "Proximal," as the study used image segmentation to identify grape pixels, as opposed to vegetation indices which other remote sensing studies employed. In this context, VSP refers to vertical shoot positioned trellis management.

Acquisition Method	**Model**	**Number of vines / Field size**	**Sensor**	**Management**	**Performance**	**Reference**
Proximal	Segmentation and linear correlation	10	RGB	VSP	R^2: 0.73	(Diago et al., 2012)
Proximal	Berry count and linear correlation	950	RGB	VSP or Split-V trellis, basal leaf removal	R^2: 0.6 – 0.73	(Nuske et al., 2014b, 2014a)
	Berry count calibrated to previous harvest yield and linear correlation	112	RGB		Error: -2.47 – 11.65%	
Proximal	3D bunch modeling and linear correlation	14	RGB	VSP	R^2: 0.778	(Herrero-Huerta et al., 2015)
Proximal	Berry count adjusted with Boolean model	84	RGB	VSP	R^2: 0.78	(Millan et al., 2018)
Proximal	CNN	40	RGB	VSP	R^2: 0.54	(Silver and Monga, 2019)
UAV (Proximal)	Segmentation and linear correlation	68	RGB	VSP	Accuracy: 22.2 – 91.7%	(Di Gennaro et al., 2019)
		32	RGB	VSP Vines with optimal	R^2: 0.82 RMSE: 0.67 kg/vine	

				conditions selected		
UAV (1 m)	Vegetative indices and linear correlation	0.032 ha	Multispectral	VSP (78%) and Umbrella (22%)	R^2: 0 – 0.71	(Carrillo et al., 2016)
UAV (0.03 m)	Vegetative indices and linear correlation	0.9 ha	Multispectral	Cordon spur-pruned	R^2: 0.33 – 0.80	(Matese and Di Gennaro, 2021)
UAV (0.07 m)	Vegetative indices and MLP model	~0.28 ha	Multispectral	VSP	R^2: 0.65 – 0.95	(Ballesteros et al., 2020)
Satellite (1000 m)	Vegetative indices and linear correlation	10,000 ha	Multispectral	Varied	R^2: 0.73 – 0.88	(Cunha et al., 2010)
Satellite (10 m)	Vegetative indices and linear correlation	1.4 ha	Multispectral	Double cross-arm	R^2: 0 – 0.33	(Anastasiou et al., 2018)
Satellite (10 m)	Vegetative indices and linear correlation	56 ha	Multispectral	Unspecified	R^2: 0.04 – 0.59	(Sun et al., 2017)

Proximal imaging conducted from the ground has seen extensive research and has demonstrated better relationships between predicted and measured yield values, as shown in Table 2.4 (Bargoti and Underwood, 2017; Gené-Mola et al., 2019; Gongal et al., 2015; Santos et al., 2020). However, while proximal imaging is advantageous due to the increased resolution of images (millimeter resolution as opposed to 10- or 30-meter resolution in remote sensing) and visibility of the harvestable region of fruits like grapes, issues with occlusion of fruit by foliage arise as proximity increases (Gongal et al., 2015; Mu et al., 2020; Nuske et al., 2014a). Occlusion poses a considerable problem, as most models have used conventional computer vision methods with geometric and color-based features to segment or count visible grapes (Table 2.4). Some studies have demonstrated benefit from incorporating variable grape visibility into yield estimation models for improving performance using statistical modeling (Millan et al., 2018; Nuske et al., 2014b). However, these corrective procedures rely on prior information about the fruit structure or past data. Nevertheless, even with the complication occlusion brings, performance of the models has been encouraging, with R^2 values as high as 0.82 for the relationship of predicted and measured yield (Di Gennaro et al., 2019). That said, previous studies have been limited in scope. Specifically, almost all studies use vines trained with a vertical shoot positioned (VSP) trellis and heavily pruned canopies (Table 2.4). This trellis type, while common in high value vineyards, is less common in the California Central Valley, where high heat and low humidity necessitate increased shading (Hayman et al., 2012). Moreover, mechanical vine management, which has become very common in the Central Valley, results in increased occlusion of grape clusters by the canopy, as mechanical trimming systems cannot selectively remove shoots obscuring grape clusters (Kaan Kurtural and Fidelibus, 2021). In addition to being largely limited to VSP trellises, the previous studies in Table 2.4 have made use of unsupervised CV and ML methods, where image features are pre-determined, and the images used to develop the model are the same images used to assess performance. More specifically, no study reported performance on a representative test set. So, while previous models may be viable in environments other than those which were specifically used for model development, the lack of a true holdout set prevents conclusive determination of the robustness of the existing approaches.

Fortunately, the accessibility and performance of DL has potential to improve the predictive performance and robustness of yield estimation models on unseen data. Recently, researchers have applied more robust convolutional approaches to the task of detecting grapes in proximal imagery. Specifically, Santos et al. compared the performance of the Mask R-CNN and YOLO models in grape detection (Santos et al., 2020). Similarly, Milella et al. used a custom CNN approach for patch-based classification of grape regions in images (Milella et al., 2019). Both studies demonstrated high performance on holdout data not used for model development, but the models were trained and evaluated only on their ability to localize grape clusters in proximal imagery, not to predict yield. This is significant, as previous research has demonstrated that visible fruit in imagery does not consistently correlate with yield (Millan et al., 2018; Nuske et al., 2014a). Notably, one previous study has demonstrated the application of a CNN regression approach for prediction of grape yield from image data, where a model was developed to extract features from an image using convolutional kernels and these features were used to predict a scalar, continuous value of grape yield from the original image data (Silver and Monga, 2019). However, unlike other recent studies (Di Gennaro et al., 2019; Millan et al., 2018; Nuske et al., 2014a), images were not collected using a scalable method, such as image acquisition from a vehicle-mounted camera. Instead, images were collected manually with a smartphone. Additionally, vines were prepared specifically for imaging, as in Diago et al. (2012), via placement of a calibration marker in each frame. Finally, extensive manual processing of the images was required, including brightness normalization and manual cropping. Nevertheless, the study does provide a proof-of-concept for utilization of DL in viticultural yield estimation.

As in the food and postharvest technology domains, the flexibility and performance of DL models will likely drive future research into viticultural yield estimation from proximal image data. Current gaps in this research include accounting for grape occlusion as well as a lack of studies demonstrating utilization of more performant DL models for yield estimation as opposed to fruit localization alone.

2.4. Conclusions

Developments which allow powerful and robust computer vision and machine learning tools to be accessed by researchers and members of industry in food and agriculture have numerous benefits. Nondestructive analysis in the food industry permits simultaneous visualization and quantification of valuable metrics throughout a time-dependent process, requiring less time and material than the corresponding destructive analysis would take. Computer vision and machine learning based models in the field of postharvest quality assessment have also seen extensive use for evaluation of a large volume of food material without wasting resources. Similarly, in the agricultural environment, models that produce accurate predictions of specialty crop yield may allow growers to optimize the quality and handling of their crops without necessitating large investments in labor or destructive measurement on valuable crops. Although nondestructive quantitative imaging techniques have shown promise in food and agricultural domains already, the recent rise in high-performance and flexible deep learning models has the potential to increase the accuracy and accessibility of these techniques. However, future research into how to best apply generic deep learning models to these specific domains is necessary to support their use on a large scale.

CHAPTER 3. NONDESTRUCTIVE CHARACTERIZATION OF STRUCTURAL CHANGES DURING IN VITRO GASTRIC DIGESTION OF APPLES USING 3D TIME-SERIES MICRO-COMPUTED TOMOGRAPHY

3.1. Abstract

An in-depth understanding of food structural breakdown during gastric digestion is paramount for development of health-promoting foods. This study presents a novel application of nondestructive time-series micro-computed tomography (micro-CT) to study structural breakdown during in vitro gastric digestion of apples (var. *Granny Smith*). Data collected from micro-CT images were compared with results from destructive analyses of apple tissue hardness (texture analysis) and moisture uptake during soaking in gastric juice or deionized water. Apples in gastric juice showed similar trends in intensity change (from micro-CT images) and hardness decrease (from texture analysis) over time compared with apples in water ($p < 0.001$). Apples in gastric juice or water exhibited similar changes in porosity and showed similar moisture uptake ($p > 0.05$). Overall, micro-CT imaging allows for assessment of changes along with detailed structural characterization of solid foods during in vitro gastric digestion.

Published as:

Olenskyj, A.G., Donis-González, I.R., Bornhorst, G.M., 2020. Nondestructive characterization of structural changes during in vitro gastric digestion of apples using 3D time-series micro-computed tomography. J. Food Eng. 267, 109692. https://doi.org/10.1016/j.jfoodeng.2019.109692

3.2. Introduction

The effect of food on human health has given rise to the study of food functionality (McClements et al., 2009). Functionality in food systems describes food that provides basic nutrition as well as desirable health effects, which can include satiety, bowel regulation, and/or sustained energy (Lähteenmäki, 2013). These health benefits are related, as they arise due to the interaction between food and the body. Therefore, to optimize the potential for positive health impact upon consumption of a food, it is vital to understand the physical and chemical mechanisms behind food behavior in the body (Grundy et al., 2016a; Mezzenga et al., 2005; Norton et al., 2014).

When considering solid food products, one major limiting factor behind the exertion of positive health effects is food structural breakdown (Bornhorst et al., 2015; Kong and Singh, 2009b). Food structural breakdown occurs as a result of physical and chemical stresses placed on the food during mastication and gastric digestion. These stresses occur over multiple length scales and include moisture uptake, chemical hydrolysis, erosion, and fracture (Kong and Singh, 2009a; Norton et al., 2014; Singh et al., 2015). Although both the physical and chemical processes which create these stresses are carried out simultaneously in the body, isolation of a subset of processes allows for a better understanding of individual aspects of digestion (Minekus et al., 2014). As these individual physical and chemical processes cannot be separated in vivo, in vitro digestion techniques can be utilized to understand the fundamental mechanisms of food breakdown during different stages of digestion (Hur et al., 2011). For example, in vitro gastric digestion can be performed such that the food does not experience mechanical stresses, allowing for isolation of the effects of moisture uptake and chemical hydrolysis.

Still, eliminating the impact of mechanical stresses using in vitro methods does not remove all complexity from the study of gastric digestion. Gastric digestion is a multiscale process, specifically with regard to food structural breakdown (Bornhorst et al., 2016; Ho et al., 2013). The length scales on which breakdown occurs during digestion are typically divided into macroscale, microscale, and nanoscale. The macroscale generally represents properties on the length scales of 10^{-3} to 1 m. Food microstructure has characteristic lengths between 10^{-7} to 10^{-3} m, and nanoscale considerations are typically between length scales of 10^{-9} to 10^{-7} m (Aguilera, 2005; Ho et al., 2013). Although food properties at these scales can be studied individually, properties at smaller length scales influence those at larger length scales, and vice versa (Ho et al., 2013; Mebatsion et al., 2008). Therefore, mechanistic studies of gastric digestion need to account for processes occurring at multiple scales (Bornhorst et al., 2016). In the context of food structural breakdown, previous attention has been focused on the macroscale (Drechsler and Ferrua, 2015; Kong and Singh, 2009b) and/or microscale (Hur et al., 2009; Kong and Singh, 2009b; Morell et al., 2017; Shelat et al., 2014; Wooster et al., 2014) changes during digestion.

Macrostructural properties and their changes during digestion have been surveyed using a variety of techniques. Texture analysis has been used to assess changes in hardness of foods subjected to in vitro gastric digestion (Kong and Singh, 2008; Mennah-Govela and Bornhorst, 2016b). Similarly, rheometry has been used to monitor bulk fluid properties of digesta during gastric digestion (Bornhorst et al., 2013; Shelat et al., 2014; Wooster et al., 2014). It should be noted that most methods of measuring macrostructural properties involve destruction of the food matrix such that a sample cannot be measured more than once during the process (M. Bourne, 2002; Chen and Opara, 2013).

Coupled with macroscale characteristics, visualization of food microstructure during digestion is possible using various forms of microscopy and tomography. Optical microscopy techniques have been used to demonstrate changes in food cellular structure as a result of gastric digestion (Chen et al., 2011; Mennah-Govela and Bornhorst, 2016b). However, similar to macrostructural methods, samples can only be assessed once due to the destructive nature of these techniques. Additionally, alteration of the food matrix may occur during sample preparation, which can involve freeze drying, embedding, and slicing of sample tissue (Dalmau et al., 2017; Llull et al., 2002; Mennah-Govela and Bornhorst, 2016b). For this reason, three-dimensional (3D) nondestructive tomographic techniques are advantageous in characterizing changes in microstructure without causing changes that may occur during sample preparation. Two of the most prevalent nondestructive techniques are magnetic resonance imaging (MRI) and X-ray micro-computed tomography (micro-CT) (Herremans et al., 2014a; Lammertyn et al., 2003).

While MRI is an effective method for studying moisture transport (Gulati et al., 2015; Xu et al., 2017), 3D MRI studies on food structure typically use voxel sizes of greater than 50 µm in-plane, with slice thicknesses on the order of millimeters (Bernin et al., 2014; Groß et al., 2017; Heyes and Clark, 2003). Furthermore, acquiring these high-resolution images requires relatively long acquisition times. For example, a recent study of cherry tomato with 55 µm in-plane resolution and 0.5 mm slice thickness required an acquisition time of over 2 hours (Groß et al., 2017). On the contrary, micro-CT studies of apple microstructure have utilized isotropic voxel sizes of below 10 µm (Herremans et al., 2013; Mendoza et al., 2007). In the context of fruit imaging, studies have shown that components of fruit microstructure, such as void spaces in apples, are on the order of 100 µm in diameter (Verboven et al., 2008; Vicent et al., 2017). Therefore, enhanced resolution capability demonstrates an advantage of micro-CT over MRI for imaging apples.

However, although obtaining this level of resolution is a major advantage of micro-CT, the time required for image acquisition is still a limiting factor. In micro-CT imaging, attenuation of X-rays is measured by a detector and 2D projections are saved as 16-bit images. The amount of attenuation, typically measured in Hounsfield units, is directly proportional to the density of the object being scanned (Schoeman et al., 2016). Rotating the sample around the z-axis between successive measurements of X-ray attenuation allows for the reconstruction of a 3D map of attenuation from the 2D projections. In this 3D image, the intensity of each voxel (volume element) represents the density of the sample in that location (Baker et al., 2012). This density map can be used to distinguish between air and liquid or to assess subtle variations in density.

Due to the iterative nature of the method, micro-CT scanning is a compromise of time, resolution, and quality (Schoeman et al., 2016). Acquisition time can be improved by using a lower resolution, which requires fewer projections. Likewise, the number of projections can be decreased without changing resolution, but the likelihood of image artifacts increases (Barrett and Keat, 2004). Some of these compromises can be overcome by using synchrotron radiation, but these experiments require access to a synchrotron radiation source (Verboven et al., 2008). Additionally, work has been done in applying iterative techniques to image reconstruction to obtain images with a lower number of projections (Willemink and Noël, 2019). However, these methods are still in exploratory stages. Nevertheless, high-resolution scans of millimeter-scale objects have been recorded over the course of an hour or even faster with modern desktop instruments (Herremans et al., 2013; Mendoza et al., 2007; Trinh et al., 2013). Studies utilizing distinct samples for successive scans have used CT methods to observe microstructural changes due to frozen storage (Ullah et al., 2014; Zhao and Takhar, 2017), frying (Alam and Takhar, 2016; Miri et al., 2006), and extended storage (Herremans et al., 2014a, 2013).

Remaining conscious of its limitations, previous researchers have even been able to utilize micro-CT methods for 3D scanning of time dependent processes in situ. This method of analysis is also referred to as four-dimensional (4D) imaging, where the fourth dimension is represented by time (Verboven et al., 2018). 4D techniques have been applied to food in the past, such as to visualize milk foam decay (Eggert et al., 2014) as well as to monitor changes in ice cream microstructure during temperature variation (Pinzer et al., 2012). 4D methods have also been applied to time-dependent food processes such as gas bubble formation and growth in bread dough (Trinh et al., 2013; Turbin-Orger et al., 2015). Of these studies, Trinh et al. utilized a desktop CT source as opposed to a synchrotron source. This limitation required the authors to use a longer imaging time of 45 minutes to maintain acceptable image quality at a voxel size of about 10 µm. Still, the selection of a desktop source allowed for the study to be performed outside of a specialized synchrotron facility (Trinh et al., 2013).

High-quality images and quantitative, interpretable data obtained by previous researchers in the past demonstrate that high-resolution micro-CT methods may be useful to characterize changes during food digestion processes, as these changes take place on time scales of minutes to hours (Drechsler and Bornhorst, 2018; Mendoza et al., 2007). As such, micro-CT is a promising method for 3D time-series imaging of in vitro gastric digestion to nondestructively visualize changes in all 3 dimensions within food microstructure that occur over a relevant time scale.

In this study, a 4D approach was utilized where the same sample was scanned in 3D over time using a desktop CT instrument to characterize changes occurring in situ. The objective was to determine if microstructural changes during gastric digestion can be detected and quantified using micro-CT. Additionally, results of the micro-CT imaging study were compared with traditional destructive approaches such as moisture content determination and analysis of tissue hardness. To gauge the applicability of micro-CT imaging for the study of digestion and its relation to these destructive analyses, apple was selected as a suitable model food due to its porous matrix (Khan and Vincent, 1993) and use in previous studies. These previous applications of micro-CT in apples have included visualization and quantification of damage (Diels et al., 2017; Herremans et al., 2013), characterization of microstructure (Herremans et al., 2014b; Mendoza et al., 2007; Vicent et al., 2017), and description and modeling of gas exchange involved in apple respiration (Ho et al., 2011; Verboven et al., 2008). Compared with those previous studies, this current study is novel for multiple reasons. First, it represents the first application of time-series micro-CT imaging to the study of gastric digestion. Additionally, the characterization of microstructural changes from image data was paired with conventional destructive analysis at each individual time point, as opposed to previous time-series micro-CT studies, which relied on analysis of image-derived microstructural properties alone during processing, and did not directly couple the image data with destructive measurements.

3.3. Materials and Methods

3.3.1. Raw Materials and Sample Preparation

Two volume bushels of apples (var. *Granny Smith*, size 56) were acquired from a local produce wholesaler (General Produce, Sacramento, CA) and stored at 1ºC for up to five weeks.

Apples with initial brix between 12 and 15ºBx were cut into 12.7 mm (½ inch) strips along the direction of the apple core using a potato cutter and further cut into cubes using a knife. Any cubes containing seeds, peel, or core were discarded. The 12.7 mm cube size was chosen to give a large region of interest for CT image analysis (Fig. 3.1A-B).

Gastric juice was formulated according to Bornhorst and Singh (2013). The media was made by dissolving 8.78 g/L of NaCl (Fisher Chemical, Pittsburgh, PA), 1.5 g/L type II mucin (Sigma Aldrich, St. Louis, MO), and 1 g/L porcine pepsin (MP Biomedicals, Santa Ana, CA) in deionized water adjusted to pH 1.8 with 3 M HCl.

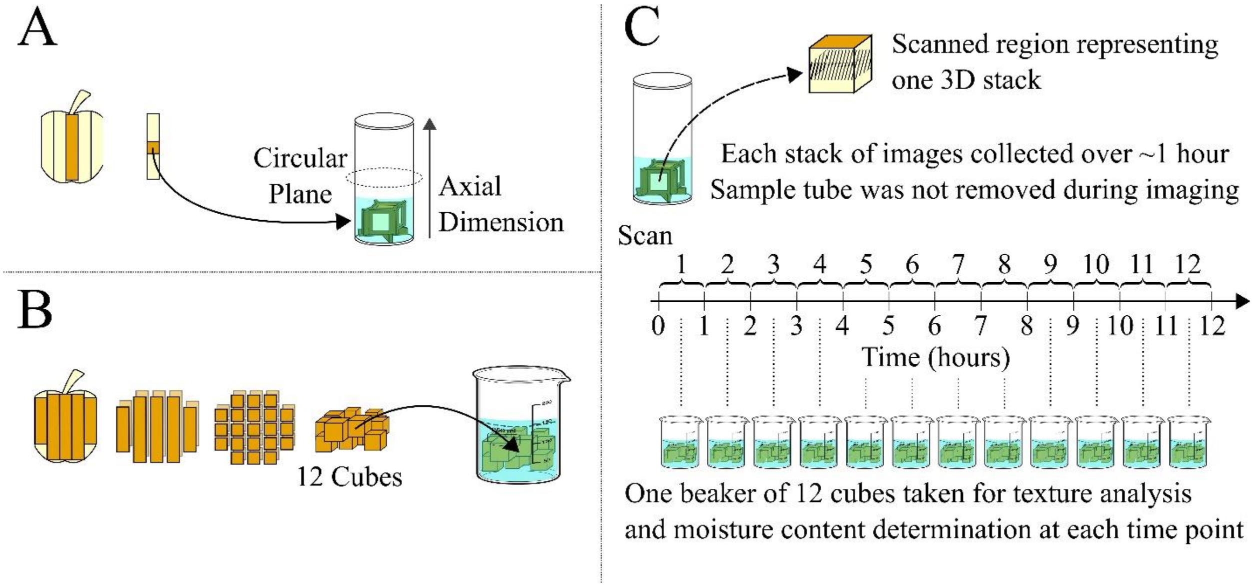

A
Circular Plane
Axial Dimension
B
12 Cubes
C
Scanned region representing one 3D stack
Each stack of images collected over ~1 hour
Sample tube was not removed during imaging
Scan
1 2 3 4 5 6 7 8 9 10 11 12
0 1 2 3 4 5 6 7 8 9 10 11 12
Time (hours)
One beaker of 12 cubes taken for texture analysis
and moisture content determination at each time point

Figure 3.1. Preparation of apple cubes for micro-CT imaging (A) and destructive testing (B). Timeline demonstrating the times at which images were recorded by the micro-CT instrument as well as times at which samples were taken for destructive testing (texture analysis and moisture content), indicated by the vertical dashed lines (C).

3.3.2. Experimental Design

Apples were incubated in deionized water (pH of approximately 5.7), simulated gastric juice, or air (used as a control). For the air treatment, cubes were wrapped in paraffin film (Bemis, Oshkosh, WI) to prevent moisture loss. For water and simulated gastric juice treatments, 8.75 mL of liquid per apple cube was used, as this was determined to be a sufficient volume to submerge a single cube in the micro-CT sample tube.

Two complementary methods were used to assess physical changes in apples during incubation: micro-CT imaging and destructive testing (Fig. 2). 3D micro-CT imaging was performed on a single apple cube continuously over a 12.6-hour incubation in either gastric juice or water. This length of time, while not physiologically comparable to human digestion, allowed for the observation of microscale changes occurring in apple tissue during incubation. Destructive testing of hardness and moisture content was also performed on apple cubes incubated with similar conditions as compared to those during micro-CT scanning (Fig. 3.1C). All trials were performed in triplicate.

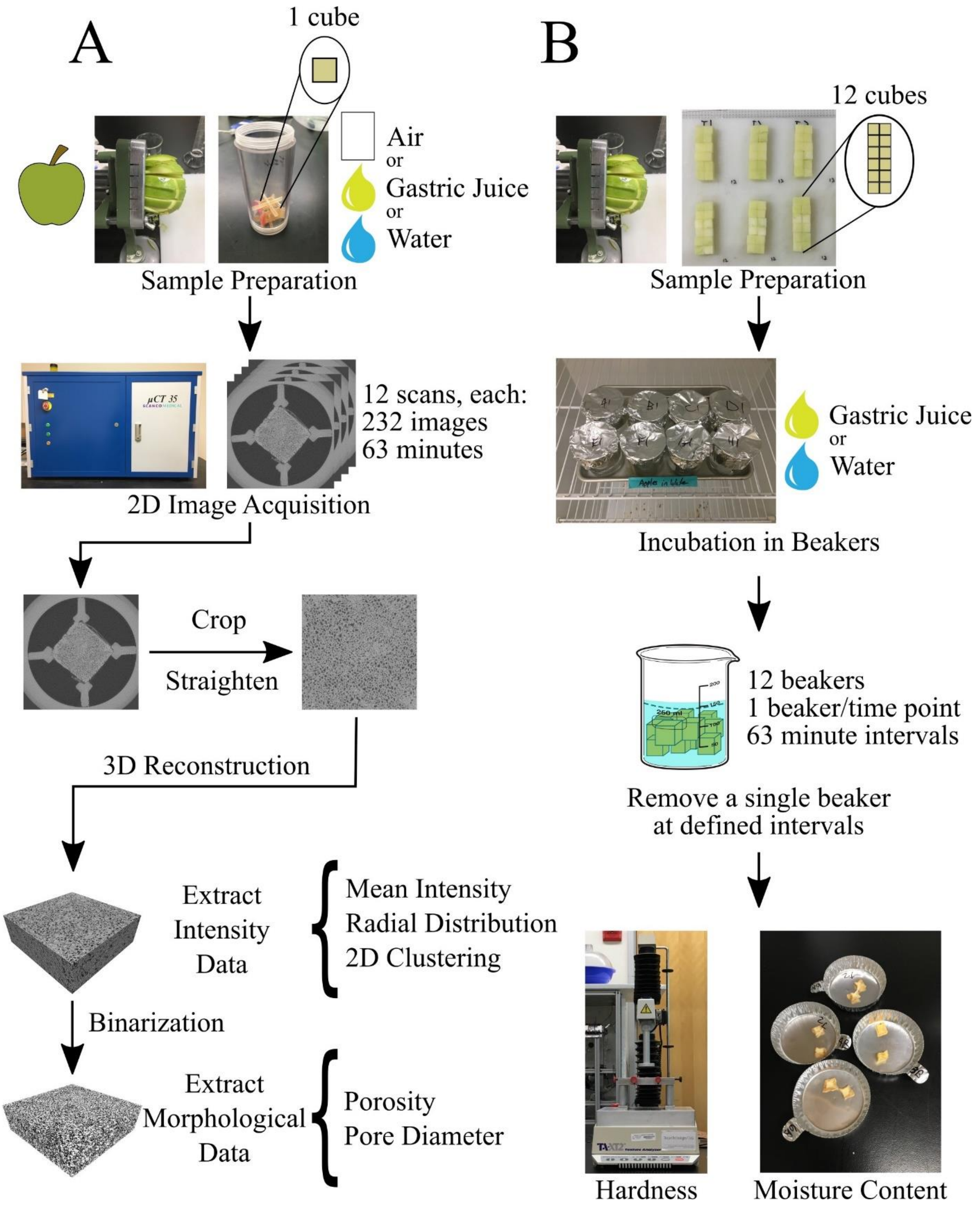


Figure 3.2. An outline of procedures used to obtain data from nondestructive micro-CT imaging (A) and destructive analysis of moisture content and hardness (B).

3.3.3. Micro-CT Imaging

3.3.3.1. Image Acquisition

For each micro-CT scanning session, one cube from an apple was selected from the outer middle region of the apple (Fig. 3.1A). Cubes were placed in a 3D printed holder such that the face of the cube which formerly faced away from the apple core (Fig. 3.1A) was facing vertically upwards in the sample tube, indicated by a dark-shaded face of the apple cube in Fig. 3.1C. The holder was used to immobilize the cube in a cylindrical plastic tube for use in the CT scanner (28.85 mm inner diameter). During micro-CT scanning, the temperature directly underneath the sample tube was logged every 2 minutes using a temperature logger (iButtonLink, Whitewater, WI).

Micro-CT scanning was performed using a Scanco uCT 35 Evaluation System (Scanco USA, Wayne, PA), which was maintained and calibrated weekly using hydroxyapatite phantoms by the UC Davis Veterinary Medicine Center. Scans were taken with a voltage of 45 kVp and an exposure of 159 µAs with 1000 projections recorded over 180º. A 0.5 mm aluminum filter was used to mitigate the effects of beam hardening. The scanned area was 2048 by 2048 pixels in the circular plane of the sample tube. 2D slices were reconstructed using an adapted cone-beam filtered backprojection algorithm (Feldkamp et al., 1984). Each reconstructed 3D image stack represented 4.3 mm in the axial dimension (Fig. 3.1A), providing 18.5 µm isotropic resolution, with a scan time of 1.05 hours per 3D stack. Data were recorded in 16-bit grayscale values.

For the air treatment, the entire apple cube was scanned. This was done by scanning three contiguous 3D stacks of images. The stacks were scanned in series, with the final concatenated 3D stack representing approximately 13 mm in the axial dimension and requiring 3.15 hours of scanning time per cube. After the first scan, the apple cube was scanned sequentially three more times, resulting in four total scans per apple cube per session. This provided an incubation time of around 12.6 hours per scanning session. The air treatment was meant to serve as a control where changes over time were expected to be minimal when compared to samples in liquid.

For the gastric juice and water treatments, the scanned region in the axial dimension was decreased to allow for an increased number of scans to be conducted within the same 12.6-hour time frame. This effectively increased the resolution in the time dimension. One 3D stack (4.3 mm axial length) was selected from the middle of the apple cube to reduce edge effects (Fig 3.1C). The region within the cube was sequentially scanned 12 times for each apple cube, with each scan requiring 1.05 hours to complete (Fig 3.1C). After each 12-hour scanning session, the sample tube was removed from the instrument. In every case, apple cubes maintained their position and needed to be manually removed from the holders.

In each scan, representing one time point within a scanning session, 232 two-dimensional (2D) images were reconstructed in DICOM format (Digital Imaging and Communications in Medicine), where the x and y axes represented the circular plane and the z axis represented the axial dimension (Fig 3.1A). 2784 reconstructed images were collected per scanning session. Replicate scanning sessions were performed on three separate days for each of the three experimental treatments for a total data set consisting of 25,056 images from nine cubes, each from a different apple.

3.3.3.2. Image Pre-Processing

Image processing was performed in MATLAB 9.5 (Mathworks Inc., Natick, MA). Images were recorded with a field of view sufficient to visualize the plastic 3D-printed holder (Fig. 3.2A).

Small regions of the holder were assessed for intensity changes over time and these changes were found to average less than 0.8% intensity increase within a scanning session. As such, no adjustments were made to the intensity or contrast of the images.

For all analyses, after manual straightening, each image was cropped to a 601 x 601 pixel (11.1 x 11.1 mm) region of interest in the center of the image such that each image represented only apple tissue and edge effects were minimized (Rizzolo et al., 2013). The images were then stacked to create a single array, stored as a MAT-file, representing a 3D reconstruction.

3.3.4. Image Analysis

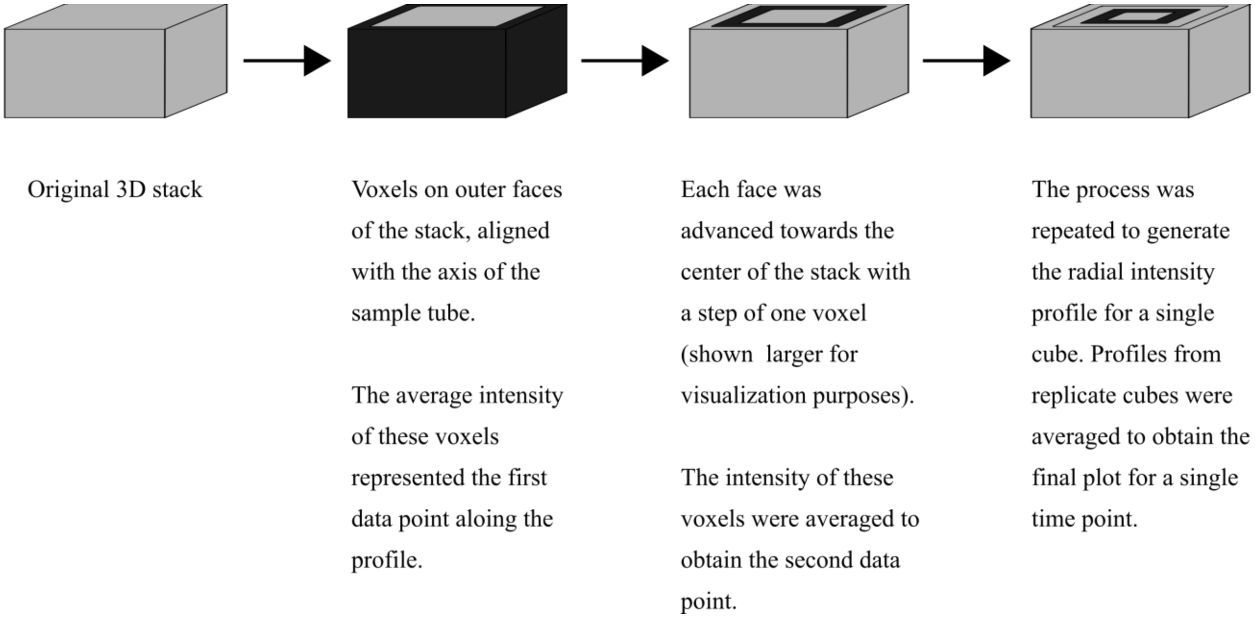
3.3.4.1. Morphological Characteristics

Morphological analysis was conducted by assessing properties of each 3D stack and noting the changes that occurred over time. Specifically, 3D stacks were binarized using Otsu's method (Diels et al., 2017; Herremans et al., 2014a, 2013; Rizzolo et al., 2013). The first scan in each scanning session of the air treatment was taken to represent unmodified apple tissue. A binarization threshold was calculated for each of these scans and the average threshold value (n = 3) was then applied to all scans in all treatments. After binarization, an opening operation with a spherical structuring element with 1 pixel radius was conducted to remove noise (Diels et al., 2017). The number of black voxels (voxel value equal to 0) over the total voxels represented the porosity, with the assumption that black voxels represented void space, and white voxels (voxel value equal to 1) represented tissue. Pore size was determined by first calculating the volume of each connected black object in the stack and solving for a diameter assuming objects were spherical (Alam and Takhar, 2016). For comparison of pore diameter values, the pore diameter threshold accounting for 50% of total volume (d_{50}) was calculated (Herremans et al., 2013). This was done by fitting the diameter to a Rosin-Rammler distribution using nonlinear least squares (Hutchings et al., 2011):

$$Q = 1 - e^{-\left(\frac{d}{d_{50}}\right)^{b} \cdot \ln(2)} \tag{3.1}$$

Where Q is the cumulative diameter (%), d is a single diameter measurement (μm), d_{50} is the median pore diameter (μm), and b is a parameter describing the distribution broadness (unitless).

3.3.4.2. Visualization and Quantification of the Digestion Process



Original 3D stack

Voxels on outer faces of the stack, aligned with the axis of the sample tube.

The average intensity of these voxels represented the first data point aloing the profile.

Each face was advanced towards the center of the stack with a step of one voxel (shown larger for visualization purposes).

The intensity of these voxels were averaged to obtain the second data point.

The process was repeated to generate the radial intensity profile for a single cube. Profiles from replicate cubes were averaged to obtain the final plot for a single time point.

Figure 3.3. Graphical representation of process used to generate radial intensity profiles for apples scanned in the micro-CT instrument. Radial profiles are shown in Figure 3.5.

Intensity variations within apple tissue were calculated using a radial intensity profile of each stack (Van Wey et al., 2014). The average intensity of the voxels on the four outer faces of the 3D stack aligned with the axial plane was taken as a single intensity value, then additional intensity values were calculated by advancing each face towards the center of the cube with a step of one voxel (Fig. 3.3). While this did not represent a true radial measurement, as the sample was cubic and not cylindrical, due to the dependence of the acquired profile on distance from the edge of the cube as opposed to location around the cube, the analysis is referred to as a radial measurement. A profile was created for each 3D stack within a scanning session and plotted such that each time point represented a single profile. Profiles within one scanning session were normalized by dividing all intensity values in the profile by the first data point, representing the outermost ring of pixels of the first 3D stack. After normalization, profiles across replicates were averaged.

To better visualize the dataset, one scanning session from each treatment was selected. 3D stacks from each scanning session were concatenated into a single stack (four stacks from an apple in air, twelve stacks from an apple in gastric juice, and twelve stacks from an apple in water). The 3D stack was then converted into a single 2D binary image using the *k-means++* clustering algorithm (Arthur and Vassilvitskii, 2007). In this implementation of k-means clustering, each (x,y) coordinate in the radial plane was taken as a distinct sample. Voxel intensities along the z-axis were treated as a set of observations corresponding to each sample. Using these values, each (x,y) coordinate was categorized into one of two groups, represented by either black (0) or white (1). These groups are intended to represent “digested” and “undigested” tissue, though this discrimination cannot be validated. Nevertheless, the operation was meant to produce an interpretable 2D binary image. In this case, the dataset consisted of one apple per treatment.

Within each scanning session, the average intensity value of each 3D stack was recorded. Although the average intensity is not a unique description of a 3D stack, the value was calculated as a digital equivalent of destructive determination of hardness, where an entire apple cube is compressed to obtain a single hardness value. Average intensity values were then normalized to the intensity of the first scan and plotted against incubation time. For these intensity values to be comparable to destructive testing (Section 3.3.5.2), intensity increases were multiplied by (-1) such that normalized intensity decreased during incubation.

3.3.4.3. Statistical Analysis of Image Data

The effect of incubation time (scan 1-12 over 12.6 hours) and experimental treatment (air, gastric juice, water) on overall intensity, porosity, and mean pore diameter was determined in SAS 9.4 (SAS Institute, Cary, NC) using a mixed model with time as a repeated measure for each scanning session. Tukey's multiple comparison test was used to compare mean values where main effects were significant. All values are reported as mean ± standard error of the mean (n = 3) unless indicated otherwise.

3.3.5. Static Incubation

3.3.5.1. Static Incubation Experimental Setup

Apples were cut into cubes as described in Section 3.3.1. Property measurements were conducted initially (before incubation), and at time points of 0.53, 1.60, 2.67, 3.73, 4.80, 5.87, 6.93, 8.00, 9.07, 10.13, 11.20, and 12.27 hours of incubation in simulated gastric juice or water. These times corresponded with the time that each of the 12 scans recorded in the micro-CT scanner were half-way completed (Fig. 3.1C). For each time point, 12 cubes were placed in a 250 mL beaker containing 105 mL (8.75 mL per cube) of either gastric juice or water (Fig 3.1B). Soft plastic mesh was placed over the cubes to keep cubes submerged in liquid during incubation in a low-temperature incubator (VWR, Radnor, PA) at 33ºC. Although this temperature is lower than typically utilized for in vitro digestion (37ºC), it was selected based on the recorded temperature within the micro-CT scanner during scanning. Preliminary trials showed that apples wrapped in paraffin film did not show significant physical property changes over 12.27 hours at 33ºC. Therefore, static incubation was not performed for the air treatment.

3.3.5.2. Physical and Chemical Property Measurements

At each time point, a single beaker was removed from incubation and the 12 cubes were removed from the liquid using a strainer. To provide information on the environment in which the apples were incubating, pH and brix of the incubation medium was determined using a digital pH meter (Fisher Scientific, Pittsburgh, PA) and digital refractometer (Hanna Instruments, Woonsocket, RI). Brix of the initial gastric juice (1.6 ºBx) was subtracted from all brix values collected in the gastric juice treatment. Moisture content of the apple tissue was determined gravimetrically by drying two apple cubes to constant mass at 70ºC under approximately 0.85 bar vacuum for 24 hours (AOAC, 2000). Hardness of the 10 remaining cubes was measured using a TA.XT2 Texture Analyzer (Texture Technologies Corp., Hamilton, MA) with a 45 mm cylindrical probe and a 50 kg load cell. Samples were compressed to 50% strain at 1 mm/s (Paula and Conti-Silva, 2014), and the peak force during compression was used as a measure of hardness (Texture Technologies Corp., Hamilton, MA). Within each replicate digestion, hardness values were normalized by dividing by the hardness of the undigested sample (Drechsler and Bornhorst, 2018).

3.3.5.3. Statistical and Data Analysis of Static Incubation Trials

Changes in hardness over incubation time were fit to a modified two-parameter Weibull model:

$$\frac{H_t}{H_0} = e^{-(kt)^{\beta}} \tag{3.2}$$

H_t represented hardness at time t in hours and H_0 signified initial hardness. The scale parameter k (h^{-1}) and shape factor β (dimensionless) were estimated from nonlinear least squares fitting. This model was selected based on its previous application to fitting of hardness decrease over digestion time (Bornhorst et al., 2015; Drechsler and Bornhorst, 2018) as well as its utilization in solid loss during digestion (Kong and Singh, 2011).

The effect of experimental treatment on Weibull model parameters was determined using a two-tailed Student's t-test. The effect of experimental treatment (simulated gastric fluid or water) and treatment time (0 – 12.27 h) on pH, brix, moisture content, and hardness was determined with a two-way completely randomized analysis of variance computed using SAS 9.4 (SAS Institute, Cary, NC). A mixed model was implemented with replicate as a random factor. All values are reported as mean ± standard error of the mean (n = 3) unless indicated otherwise.

3.4. Results and Discussion

3.4.1. Image Acquisition and Analysis

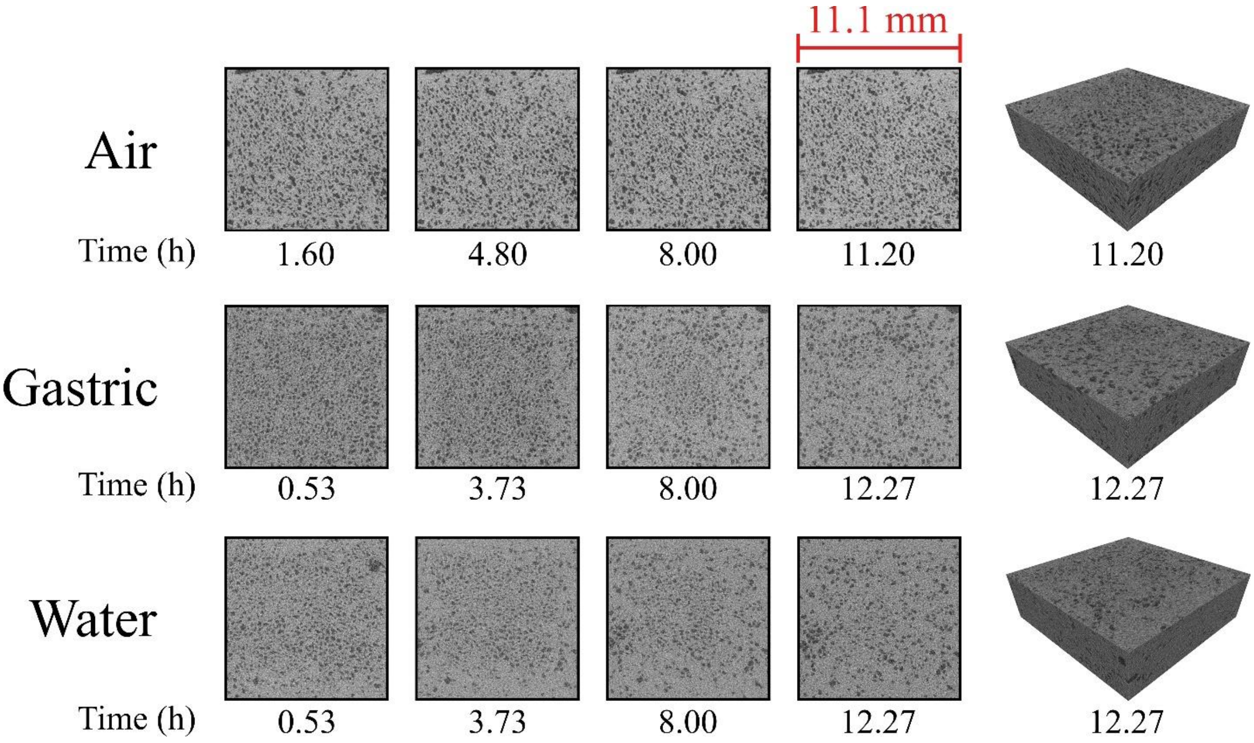


Figure 3.4. Reconstructed images taken from the center of each 3D stack. 2D images are displayed without contrast or intensity adjustment. 3D renderings from the final scan in each treatment have been histogram equalized for clarity.

Prior to conducting this study, preliminary trials (results not shown) demonstrated that a 1-hour acquisition time with 1000 projections provided adequate resolution to visualize differences over the scan time selected. With these parameters, artifacts like blurring or shifting of the tissue during incubation were not observed. Reconstructions and 3D renderings for each treatment are included in Fig. 3.4. Although 12.6 hours is a longer time than typically studied in the context of digestion, results demonstrate that the reduced temperature of 33ºC in this system slowed the digestion process to allow for appropriate comparisons to be made with a shorter experiment conducted at 37 ºC. Based on a previous study, *Granny Smith* apples digested for 4 hours at 37ºC decreased to 28% of their initial hardness (Olenskyj et al., 2020). This value is very similar to the value obtained in this study (27%) after 8 hours of digestion at 33ºC. Therefore, the increased image acquisition time was accounted for by the reduced digestion time caused by the lower temperature, and high-quality time-series images could be captured.

3.4.1.1. Apple Morphological Characteristics

Apple porosity was calculated from 3D image stacks for each scan (Fig. 3.5). Incubation time was the only main effect shown to have a significant effect on porosity (F-value = 23.81, $p <$ 0.001). However, the interaction between treatment and incubation time was found to be significant (F-value = 2.31, p = 0.0156).

Porosity of apples in air was found to be 21.99 ± 1.57% for the first scan and did not change significantly over the scanning time as assessed by post-hoc analysis (p > 0.05). Due to the larger voxel size used in this study as compared with previous studies on apple porosity, some error may be expected in the determined porosity value (Mendoza et al., 2007). However, as the objective of this study was to quantifying changes over time and between treatments, the resolution was sufficient to resolve these differences. Any error resulting in the voxel size would be present in all measurements, but still allows for comparisons to be made directly between different data sets collected in the current study. Moreover, two previous studies implementing micro-CT imaging on apples determined the porosity of *Jonagold* apples to

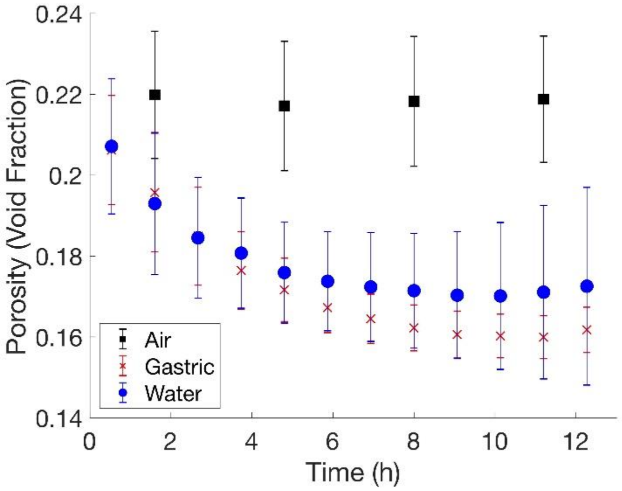


Figure 3.5. Porosity of apples incubated in air (■), water (●), or gastric juice (x) over incubation time calculated from micro-CT stacks. Error bars represent standard error of the mean (n = 3).

be 23%, which is similar to the result obtained in this study (Verboven et al., 2008; Vicent et al., 2017).

Apples in both gastric juice and water showed a decreasing porosity over incubation time ($p < 0.001$) from an initial porosity of 20.62 ± 1.35% for apples in gastric juice and 20.71 ± 1.67% for apples in water. Porosity for apples in gastric juice and water after 12.27 hours of incubation was 16.18 ± 0.56% and 17.26 ± 2.44% respectively. Although apples in liquid media showed significant porosity decreases over time, neither of the liquid treatments was significantly different from apples in air at any time point ($p > 0.05$). This was likely due to the large variation seen in porosity of the apples in air. The gradual reduction in porosity for the apples incubated in gastric juice or water was likely due to moisture filling void spaces in the tissue. A similar phenomenon was observed in a micro-CT and MRI study of watercore disorder in apple tissue, in which free water filled voids, reducing the measured porosity (Herremans et al., 2014a). The lower initial porosity for apples in liquid relative to apples in air was likely due to moisture influx, which was occurring during the first scan. The lack of significant differences between gastric juice and water treatments ($p > 0.05$ at all time points) suggests similar moisture uptake for both treatments, which was confirmed using destructive measurements (Section 3.4.2.1).

Pore diameter (d_{50}) was calculated along with porosity for each scan. Overall statistical analysis showed treatment, incubation time, and their interaction exerted significant effects on pore diameter (F-values = 711.39, 2.66, and 6.47, respectively; $p < 0.001$ for the treatment and interaction effect and $p = 0.0092$ for the time effect). Median pore diameter for the central stack in the axial dimension of apples in air was 173.53 ± 3.98 μm for the first scan and 181.20 ± 4.71 μm for the final scan. This difference was found to be significant by Tukey's multiple comparison ($p < 0.0001$). This may have been due to a small amount of moisture migration out of the apple into the space between the apple tissue and the inside surface of the paraffin film. Initial pore diameter for apples in gastric juice and water was 65.40 ± 1.50 µm and 64.19 ± 2.34 µm, respectively. The smaller pore diameter may have been due to swelling of the tissue in the aqueous media. Pore diameter was not found to increase significantly over time for either gastric or water treatments ($p > 0.05$).

Previously calculated average pore diameter of apples has been similar to those found here for *Granny Smith* apples in air, with 220 and 100 µm pore diameters found for *Braeburn* and *Verde doncella* apples, respectively (Herremans et al., 2014a, 2013). The variation in pore diameter from the previous studies compared to the current study may be because apples have different morphological characteristics depending on their variety (Vincent, 1989).

3.4.1.2. Visualization of Changes to Apple Tissue during the Digestion Process

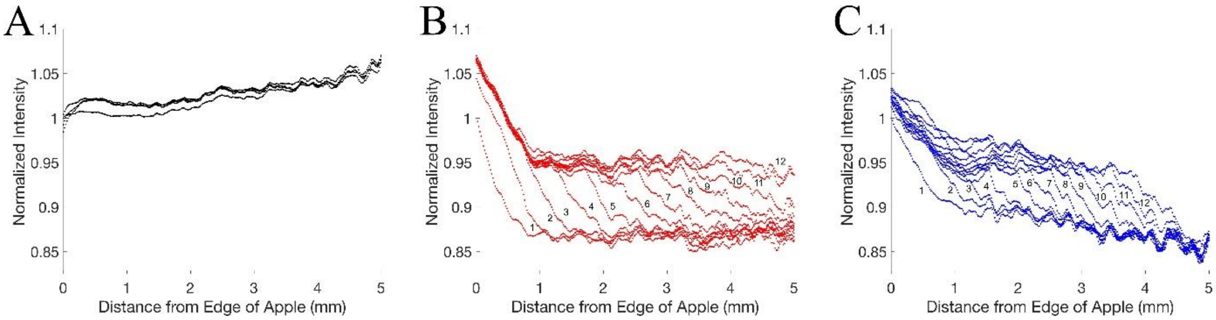


Figure 3.6. Radial intensity profiles along the x-y plane over incubation time for apples in (A) air, (B) gastric juice, and (C) water. Intensity profiles represent intensity values averaged over the z axis and averaged again over all biological replicates, then divided by the first intensity value of the first profile such that all sets start from a normalized value of 1. Numbers adjacent to curves represent the scan number in the scanning session from which the curve is derived.

The nondestructive nature of CT imaging allowed for differences in radial intensity profile over time to be visualized (Fig. 3.6). These radial profiles also give insight into the underlying mechanisms behind diffusion and reaction of gastric juice and water with apple tissue. For the air treatment (Fig. 3.6A), minimal change over time shows a lack of moisture migration or reaction in the samples. The overall lack of changes for apples in air during incubation also suggests that any changes seen in the gastric juice and water treatments were due to the action of the incubation liquid itself, and not the micro-CT scanner or changes in the apple tissue that occurred during scanning.

For the gastric juice treatment (Fig. 3.6B), radial profiles along the first 1 mm from the outside of the cube have high intensity, perhaps due to some sample disruption due to cutting the edge of the apple tissue. However, past this 1 mm boundary, radial profiles show curves characteristic of Type II diffusion (Hopkinson et al., 1997), where gradual penetration of gastric juice causes an increase in normalized intensity from approximately 0.87 in unmodified apple tissue to approximately 0.95 after the tissue is fully saturated with gastric juice. Further support for the mechanism of Type II diffusion can be found in the measured moisture content over time (Section 3.4.2.1), which showed a linear increase in g water/g dry matter over the incubation time (Peterlin, 1965). This behavior suggests that diffusion of solutes in digested tissue may occur much faster than in undigested tissue. Radial profiles for apples in water (Fig. 3.6C) do not display as clear of a trend, with a gradual overall decrease in intensity over the profile for all scans, even for unmodified apple tissue. However, these data demonstrate potential for nondestructive imaging methods to be used for modeling of diffusion processes during digestion in future studies.

In addition to the radial profiles, binarized image stacks, shown as 2-D representations of the incubation process in air, gastric juice and water, are displayed in Fig. 3.7. These images support the radial intensity plots (Fig. 3.6), with apples in air showing little change, apples in gastric juice demonstrating high density tissue in the center by the end of incubation (shown by white pixels), and apples in water showing a region in the center of the cube which remained unchanged (in black). Both the radial profiles and binary images indicate that the digestion of apples is diffusion-limited, with outer regions of apple tissue showing changes before inner regions. This phenomenon is supported by previous work on visualization of digested almonds, where the outer layers of tissue are affected first by the digestion medium (Ellis et al., 2004; Mandalari et al., 2008). Similarly, carrots and cheese in dyed acidic water demonstrate a front of dye penetrating the tissue from the outer edges inwards (Kong and Singh, 2009b; Van Wey et al., 2014), highlighting the diffusion-limited mechanism. Although these results provide mechanistic insight on the digestion of apples, future work in this area is needed to model this process and to determine the impact of digestion on the physical structure of the tissue. Future studies could also examine the properties of food that exert the most influence over the magnitude of diffusion within digested tissue, as well as examine the impact of structural changes that occur during digestion on diffusion of solutes and enzymes into food matrices.

3.4.1.3.Quantification of Overall Changes in Apple Tissue during the Digestion Process

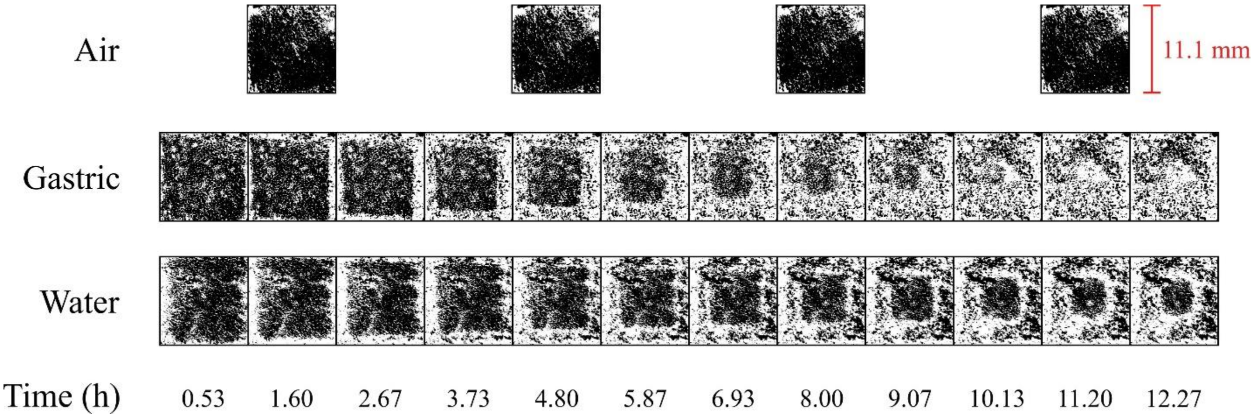


Figure 3.7. Binarized 2D images representing k-means clustering of 3D stacks collected during incubation. White pixels represent brighter (higher density) regions of the original stacks and black pixels represent darker (lower density) regions. Images shown are from one replicate of each treatment as representative examples. The listed times are the times at which each scan was 50% complete. Borders around each 2D image from each time point were added after clustering.

One of the key advantages of micro-CT image acquisition in the context of digestion is the ability to nondestructively probe the local distribution of densities as well as changing density over time within a food sample (Schoeman et al., 2016). However, in addition to requiring unique samples for each measurement, destructive measurements like analysis of hardness and moisture content assess macroscale properties, where the entire food sample contributes to a single numerical value (Ho et al., 2013; Mebatsion et al., 2008). For this reason, the 3D images collected in this study were also analyzed in a way that would be comparable to a measurement of moisture or texture by assessing the average intensity of the 3D stacks over time (Fig. 3.8).

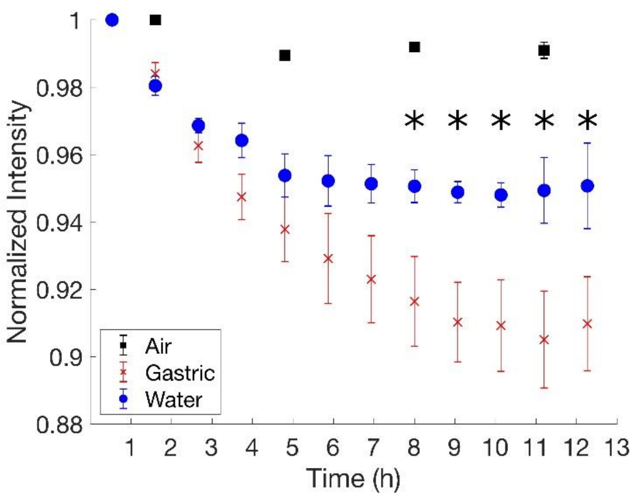


Figure 3.8. Intensity of each 3D stack for apples incubated in air (■), water (●), or gastric juice (x) over time, expressed as averages ± standard error of the mean. Intensity values are normalized to the intensity of the first stack in each scanning session of each treatment. Stars represent significant differences between values for gastric juice and water treatments (p < 0.05).

Raw data shown in Table 3.1 show that the 3D images demonstrated an increase in average intensity over time. Due to the relationship between intensity and density inherent in CT images (Lim and Barigou, 2004), it was hypothesized that data collected from micro-CT images would correlate best with moisture uptake during incubation. However, unlike the moisture uptake, the increase in intensity was not a linear increase, and the trend was not equivalent for both gastric juice and water. For this reason, the data were normalized to the initial intensity value of each scanning session (similar to how the hardness data were reported) and reported as decreases for the purpose of visualization. When the data are presented in this manner, the trend in intensity change is similar to the trend seen in hardness (Section 3.4.2.1), where all apples show nonlinear changes over the incubation time, and apples in gastric juice show a greater degree of change as compared with apples in water. This relationship suggests that either destructive testing of hardness during gastric digestion or nondestructive assessment of intensity change using micro-CT may be able to characterize overall changes to tissue during gastric digestion.

Table 3.1. Average intensity of 3D stacks collected from the micro-CT instrument over time for all treatments (n = 3). Intensity values were recorded as 16-bit signed integers (range of -32,768 to 32,767).

Time (h)	Average Intensity (CT units)		
	Air	Gastric	Water
0.53	--	2861 ± 53.7	2858 ± 64
1.6	2796 ± 30.3	2907 ± 60.5	2915 ± 70
2.67	--	2968 ± 47.9	2948 ± 61
3.73	--	3011 ± 37.6	2960 ± 55.1
4.8	2825 ± 32.2	3038 ± 32.4	2990 ± 49.5
5.87	--	3063 ± 21.4	2994 ± 48.3
6.93	--	3080 ± 21.4	2997 ± 55.6
8	2818 ± 32.5	3099 ± 19.9	2999 ± 53.8
9.07	--	3117 ± 24.8	3004 ± 59.6
10.13	--	3120 ± 19.8	3007 ± 73.6
11.2	2821 ± 28.4	3131 ± 17.4	3004 ± 89.5
12.27	--	3118 ± 18.2	3001 ± 97.5

Overall statistical analysis showed treatment, incubation time, and their interaction exerted significant effects on intensity (F-values = 77.66, 49.53, and 8.11, respectively; $p < 0.001$ for all effects). Under post-hoc analysis, apples in air showed no significant change in intensity throughout the incubation. However, 3D stacks of apples in water and gastric juice both showed significant intensity difference from initial intensity after 2.67 hours ($p < 0.05$). Starting from a normalized intensity of 1.00, the normalized intensity of apples in gastric juice and water decreased to 0.963 ± 0.0050 and 0.969 ± 0.0021 respectively after 2.67 hours. After 4.8 hours of incubation, the gastric juice and water samples were significantly different than the air samples ($p < 0.001$). Apples in air remained at an intensity of 0.989 ± 0.0012, and normalized intensity of apples in gastric juice and water decreased to 0.938 ± 0.0096 and 0.954 ± 0.0064 after 4.8 hours, respectively. After 8 hours of incubation, the intensity values from gastric juice and water treatments were significantly different from each other ($p < 0.05$), with apples in gastric juice showing a greater intensity change at all incubation times from 8 – 12 hrs. At the end of the incubation, apples in gastric juice had an intensity of 0.910 ± 0.014, whereas normalized intensity for apples in water was 0.951 ± 0.013.

This greater intensity change over the incubation time for apples in gastric juice relative to apples in water contrasted with the porosity changes over time in both treatments, i.e. porosity of apples in gastric juice and water were similar throughout the incubation (Section 3.4.1.1). The discrepancy between these two measurements may be due to the binary nature of the porosity measurement. When binarizing an image, any given voxel is either tissue (1) or void (0). With similar moisture uptake, a lack of significant difference between the two treatments is expected. However, measurement of mean intensity provides additional information, wherein microstructural changes due to prolonged contact with gastric juice may have been reflected by an increase in intensity of a given voxel over the incubation time. This increase in intensity is not captured in the porosity measurement, as the binarized voxel would be represented by a (1) regardless of the raw intensity value.

3.4.2. In Vitro Incubation and Destructive Physical Property Measurement

3.4.2.1. Quantification of Hardness and Moisture Changes in Apple Tissue during Incubation

Overall statistical analysis showed treatment, incubation time, and their interaction exerted significant effects on hardness (F-values = 578.09, 521.52, and 46.48, respectively; $p < 0.001$ for all effects). Apple initial hardness was 70.6 ± 0.48 N. Hardness decreased significantly over incubation time for both treatments ($p < 0.001$), though apples incubated in water had a significantly higher hardness averaged over all time points ($p < 0.001$). The interaction between incubation liquid and incubation time was significant ($p < 0.001$), demonstrating that the incubation liquid affected the change in hardness over time. Parameters from the modified Weibull model, k (h^{-1}) and β (dimensionless) were 0.148 ± 0.00503 h^{-1} and 1.09 ± 0.110 for apples incubated in gastric juice and 0.0356 ± 0.0097 h^{-1} and 0.486 ± 0.0680 for apples incubated in water, respectively. Curve fits were found to be in good accordance with experimental data (Fig. 3.9A), with R^2 values of 0.981 ± 0.010 and 0.957 ± 0.025 for gastric juice and water, respectively. Both model parameters were significantly different between gastric juice and water ($p < 0.01$).

Hardness curves (Fig. 3.9A) demonstrate that hardness of apples incubated in water decreases at a similar rate ($p > 0.05$) to apples in gastric juice for the first 3.73 hours of incubation, with an average hardness of 45.6 ± 4.3 N for both treatments after 3.73 h incubation. However, at times greater than 3.73 h, apples incubated in water had significantly higher ($p < 0.001$) hardness compared to apples in gastric juice. After 12.27 hours of incubation, apples incubated in water had a hardness of 37.3 ± 7.9 N compared to those incubated in gastric juice, which had a

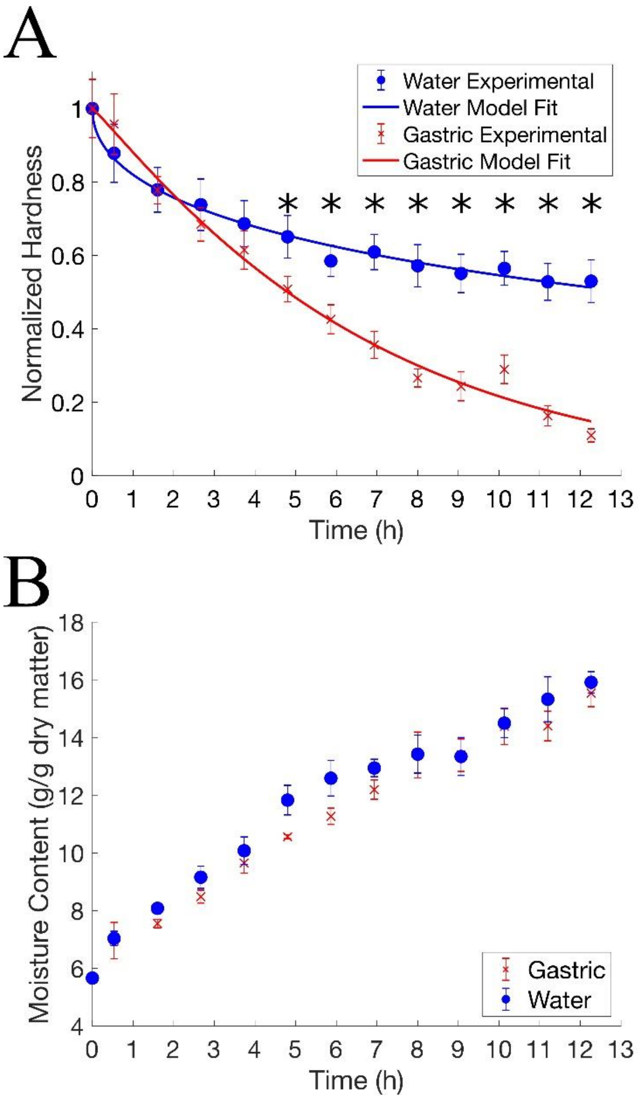


Figure 3.9. Measured normalized hardness of apple cubes in water (●) or gastric juice (x) with modified Weibull model fits (solid lines) (A) and measured moisture content (points) over incubation time (B). Error bars in both plots represent standard error of three biological replicates for each treatment. Stars represent significant differences between values for gastric juice and water treatments (p < 0.05).

hardness of 7.74 ± 4.1 N. These data align well with the mean intensity results (Section 3.4.1.3), where apples in gastric juice showed a greater degree of change throughout the incubation relative to apples in water. This relationship between image and hardness data suggests that the overall intensity of the 3D images can be used as an indicator of sample hardness in the case of *Granny Smith* apples.

Treatment and incubation time exerted significant effects on moisture content (F-values = 15.81 and 423.48, respectively; $p < 0.001$ for both effects). The interaction between main effects was not found to be significant. Interestingly, although the overall effect of incubation medium was significant, the lack of significance in the interaction between treatment and incubation time suggests that although apples in water generally had more moisture, the treatment did not influence the change in moisture content over the incubation time. The average moisture content of apples from both treatments before incubation was 5.66 ± 0.52 g water/g dry matter and increased to an average of 15.7 ± 0.8 g water/g dry matter after incubation for both treatments (n = 6; Fig. 3.9B). The lack of differences in moisture uptake between gastric juice and water treatments aligns well with porosity changes (Section 3.4.1.1), in which a lack of significant differences between treatments over time was also observed.

The overall trends of hardness decrease and moisture increase during incubation in simulated gastric fluids are similar to previous digestion studies. Previous work on digestion of apples has demonstrated that raw apples increase in moisture content during digestion (Dalmau et al., 2017). Additionally, hardness decrease of apples in the current study was similar to other carbohydrate-rich foods, where hardness was seen to decrease over time when incubated in gastric juice (Drechsler and Bornhorst, 2018). The differences observed between hardness of apples in water compared to gastric juice aligns with previous work on carrots, where carrots incubated in acidic water for one hour showed lower hardness compared to carrots incubated in neutral water after only one hour (Kong and Singh, 2009b).

Weibull model parameters k and β for apples incubated in water were both less than those of apples incubated in gastric juice. This means that hardness of apples incubated in water decreased slower (k) but with greater curvature (β) than apples incubated in gastric juice. Compared with values from previous work on modeling food hardness during digestion, the k value determined in this study for apples incubated in gastric juice, 0.148 h^{-1}, is similar to that of couscous (0.12 h^{-1}) (Drechsler and Bornhorst, 2018). Previous work has shown that *Granny Smith* apples digested in gastric juice at 37ºC for four hours had a k value of 0.339 h^{-1} and a β value of 0.8 (Olenskyj et al., 2017). This β value can be compared to a value of 1.09 found in this study with 12.27 hours of incubation at 33ºC. Comparing both k and β values obtained at 33 and 37ºC, there appears to be a marked effect of lowering temperature. Specifically, apples digested in gastric juice at 37ºC decrease in hardness faster and with greater curvature relative to apples at

33ºC. In this study, the lower temperature was used due to a limitation of the CT instrument. The instrument was not temperature-controlled, and temperature profiles taken during scanning showed a temperature plateau at 33ºC. Therefore, 33ºC was chosen for benchtop analyses. Nevertheless, the assumption in this study is that the mechanisms behind the breakdown of food structure are conserved and are merely slowed down by the reduced temperature.

These moisture uptake and hardness results, along with the non-destructive imaging analyses, suggest that decrease in hardness of apple tissue during gastric digestion is not only due to water uptake; incubation in gastric juice resulted in significantly lower hardness and greater intensity change compared to soaking in water alone. Specifically, the discrepancy between gastric juice and water samples suggest an effect of acid and/or enzyme was responsible for the differences between treatments. Typically, gastric secretions serve to break down food through the proteolytic pepsin enzyme as well as the low pH environment, which can promote pepsin activity (Kondjoyan et al., 2015; Piper and Fenton, 1965; Pletschke et al., 1995) and lead to further hydrolysis and degradation of food material (Bornhorst and Singh, 2014). Due to the limited amount of protein found in apples (0.26% according to the National Agricultural Library (2018)), enzymatic hydrolysis by pepsin was likely not the dominant mechanism in this system. Instead, this difference between gastric juice and water was likely due to pectin solubilization in the gastric juice samples due to the low pH. This mechanism has been shown to alter tissue microstructure in digestion of carrots and almonds (Kong and Singh, 2009b, 2009a). Pectin is a large component of apple tissue, comprising 10-15% of apple pomace (Herbstreith & Fox, 2000), and pectin from apple pomace is typically extracted using high heat and HCl solutions at a pH similar to the pH of gastric juice used in this study (approximately pH 1.5) (Wang et al., 2007).

This supports the thought that pectin solubilization may be the cause of the difference in hardness at the end of incubation in gastric juice compared to incubation in water.

3.4.2.2. Incubation Medium Brix and pH Changes during Incubation

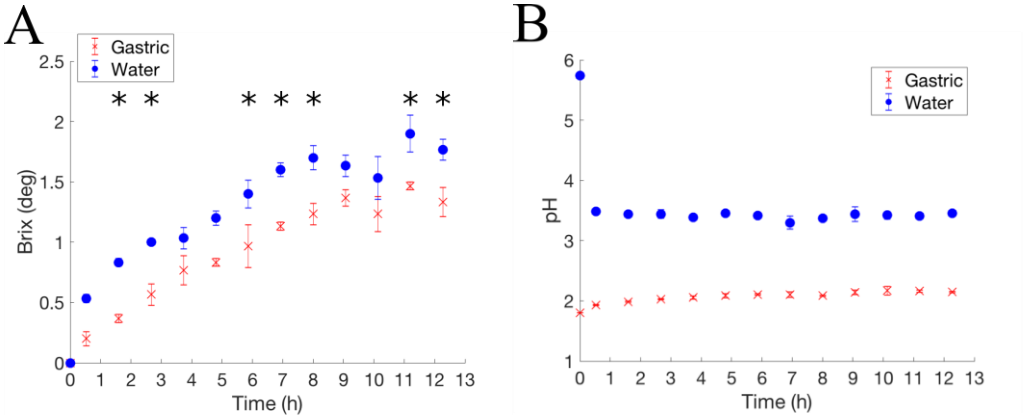


*Figure 3.10. A) Brix of incubation medium over incubation time in water (●) or gastric juice (**x**). Stars represent significant differences (p < 0.05) between values for gastric juice and water treatments. B) pH of incubation medium over incubation time in water (●) or gastric juice (**x**). Significant differences (p < 0.001) were observed at each time point. Error bars in each plot represent S.E. of three biological replicates for each treatment.*

Overall statistical analysis showed that treatment and incubation time influenced the brix of the incubation medium (F-values = 140.04 and 86.98, respectively; p < 0.001 for both effects). The interaction between the main effects was not found to be significant. Both treatments showed brix increase over the incubation time. After subtracting the brix of the gastric juice before incubation (1.6 ºBx), gastric juice samples increased from an initial value of 0 to 1.3 ± 0.1 ºBx after 12.27 hours. Water samples increased from 0 to 1.77 ± 0.1ºBx (Fig. 3.10A). Interestingly, although image analysis and tissue hardness both showed an increased influence of gastric juice on apple samples relative to water, brix of the incubation medium in the water treatment was significantly higher than that of the gastric juice treatment (p < 0.001). The higher brix in the water treatment may have been due to the decreased osmotic pressure of water relative to the gastric juice. However, the effect of osmotic pressure was likely small and limited to increased brix change in the surrounding media of the water treatment. Based on the amount of salt in the gastric juice, (8.78 g/L or 0.3 M total ions) both the gastric juice and deionized water media were hypotonic relative to the apple, which requires 0.65 M total ions in solution for isotonic media (Oey et al., 2006). As such, significant influence of osmotic pressure on apple hardness was unlikely.

pH of the incubation medium was significantly affected by treatment, time, and their interaction (F-values = 8763.94, 92.69, and 151.96, respectively; $p < 0.001$ for all effects). For samples incubated in gastric juice, pH of the incubation medium increased over incubation time from 1.8 to 2.14 ± 0.0087 after 12.27 hours, while pH of the deionized water decreased from 5.7 to 3.46 ± 0.014 (Fig. 3.10B). The magnitude of the difference in pH between gastric juice and water for the entirety of the incubation time supports a conclusion that acid concentration may have been responsible for the discrepancies between the two samples, considering apples in gastric juice experienced a significantly lower pH environment compared to apples in water.

3.5. Conclusion

X-ray micro-CT imaging allows for the nondestructive characterization of gastric digestion of apple tissue. When analyzed directly, changes shown by micro-CT demonstrate trends similar to those observed in destructive analysis of apple tissue, including hardness and moisture content. Apples in gastric juice showed both significantly increased softening (measured destructively) and significantly different mean intensity (measured nondestructively) when compared with apples in water. In addition, moisture content increase (measured destructively) and porosity decrease (measured nondestructively) over incubation were both not significantly affected by incubation treatments. These suggest that moisture influx is likely not the driving force for changes seen in texture and micro-CT images and that the differences in hardness and image analysis results between treatments are due to structural breakdown of the apple tissue in gastric juice.

Overall, this study has demonstrated the effectiveness of X-ray micro-CT imaging for assessing changes to a porous, high-moisture food matrix during simulated gastric digestion. Future considerations include improving the spatial and time resolution with an alternate imaging method such as synchrotron radiation micro-CT scanning or magnetic resonance imaging. These improvements may allow for finer or more frequent non-invasive visualization of changes to determine digestion kinetics on the microscale for in-depth computational simulation of gastric digestion.

CHAPTER 4. END-TO-END PREDICTION OF UNIAXIAL COMPRESSION PROFILES OF APPLES DURING IN VITRO DIGESTION USING TIME-SERIES MICRO-CT IMAGING AND DEEP LEARNING

4.1. Abstract

Machine learning is a promising technique to develop models, which extract relevant information from image data. This study applies convolutional neural networks trained end-to-end to predict the mechanical properties of apples (var. *Granny Smith*) from micro-CT image data collected during in vitro gastric digestion. Models were trained to directly output compression curves, allowing for representation of complex curve shapes, which changed throughout the digestion process. Models evaluated using 3-fold cross-validation demonstrated high predictive performance, with RMSE of 4.36 N and R^2 of 0.939 compared to measured data. This performance was decreased to an RMSE of 14.3 N and R^2 of 0.296 when applied to an out-of-distribution dataset. Saliency mapping used to interpret model output demonstrated a mechanistic link between typical biophysical tissue changes and model attention. Overall, the end-to-end deep learning approach represents a promising method for rapid, nondestructive evaluation of mechanical properties during food processing and digestion.

4.2. Introduction

Structural and mechanical features of food materials are known to influence their functional properties, such as shelf-life and nutritional benefits (Michel and Sagalowicz, 2008). Although microscopy and other imaging methods allow for visualization and analysis of structure using several metrics, such as porosity, cell wall thickness, and cell counting (Verboven et al., 2018), image properties are not easily mapped to mechanical properties of the bulk solid. However, techniques to assess food structure and mechanical properties, including microscopy, stress-strain testing, and rheometry, typically involve sample destruction (M. Bourne, 2002; Kaláb et al., 1995; Tabilo-Munizaga and Barbosa-Cánovas, 2004). The nature of destructive methods prevents repeated analysis on the same sample, which introduces between-sample variability and requires extensive sample preparation and large quantities of raw material (Chen and Opara, 2013; Mennah-Govela and Bornhorst, 2016b; Minekus et al., 2014; Opazo-Navarrete et al., 2018). Therefore, noninvasively assessing structural and mechanical properties is important in processes such as food digestion and drying, where changes in both attributes are significant (Bornhorst and Singh, 2014; Wang and Martynenko, 2016).

To overcome some of the previously stated limitations, tomographic imaging techniques such as micro-computed tomography (micro-CT) and magnetic resonance imaging (MRI) can be implemented (Verboven et al., 2018). Micro-CT and MRI have the advantages of generating 3-dimensional (3D) representations of food structure without significant tissue destruction, even allowing for nondestructive time-series analysis (Eggert et al., 2014; Olenskyj et al., 2020; Turbin-Orger et al., 2015). However, tomographic methods are limited in their ability to represent mechanical properties of solid foods, as the relationship between visible microstructure and mechanical properties is complex and typically non-linear (Aguilera and Lillford, 2008; Li et al., 2019).

Recent advances in machine learning techniques, such as convolutional neural networks (CNNs), have allowed for the generation of models that directly map varied, complex image data in biological systems to independently measured biophysical attributes. For example, machine learning techniques including CNNs have been applied to microstructural images of plant and animal tissue, to classify images according to disease state (Hamidinekoo et al., 2018), or segment images into tissue types (Earles et al., 2018; Théroux-Rancourt et al., 2020). These advances have allowed for rapid, and accurate analysis of biological samples, which would have otherwise required extensive time or resource investment. Recent research has also demonstrated the ability of CNNs to relate images of inorganic material, such as heterogeneous rock and simulated composites, to a continuous mechanical property value, including elastic modulus and Poisson's ratio (Li et al., 2019; Yang et al., 2018; Ye et al., 2019). In these studies, predictions made on unseen data required a relationship to be mapped between the pixels of an image to a measurement of a continuous scalar quantity. In addition to studies involving scalar prediction from 2D or 3D samples, Herriot and Spear (2020) predicted 2D maps of effective yield strength using a 3D CNN, demonstrating the flexibility of the CNN approach.

Extending on the progress made in mechanical measurement of inorganic samples, biological materials are often complex material composites, and describing them mechanically can be more precisely done by examining an entire stress-strain curve as opposed to a single scalar value. Previous research has demonstrated the ability of CNNs to predict stress-strain curves from microscopy images of collagenous tissue (Liang et al., 2017). However, as opposed to end-to-end model development, destructive curves were decomposed with principal component analysis (PCA), then these curve fit parameters were predicted using a CNN trained with a combination of unsupervised and supervised learning. Although this approach led to high performance within the tested dataset, prediction of fit parameters limits a model's robustness on samples not described by the underlying relationship of the fitted model.

Building on previous research in predicting mechanical properties from image data, it was hypothesized that CNNs may be able to directly relate food microstructure and mechanical properties. To demonstrate this technique, time-series 3D micro-CT images during in vitro digestion were used to predict independently measured uniaxial compression curves. The novelty of the present study lies in both the application of nondestructive mechanical property prediction in solid foods during in vitro gastric digestion as well as the unconstrained nature of the modeling approach. As opposed to prediction of a scalar value, such as elastic modulus or hardness, or a set of fit parameters for a curve, the tomographic images were used to predict entire apple tissue compression curves at varying digestion times and incubation conditions (simulated gastric juice and deionized water) via an end-to-end approach. The lack of constraint to a scalar value or set of parameters within a curve was hypothesized to permit the model to predict entire curves of nonspecific shapes, leading to more accurate predictions which have increased utility over scalar values alone.

4.3. Materials and Methods

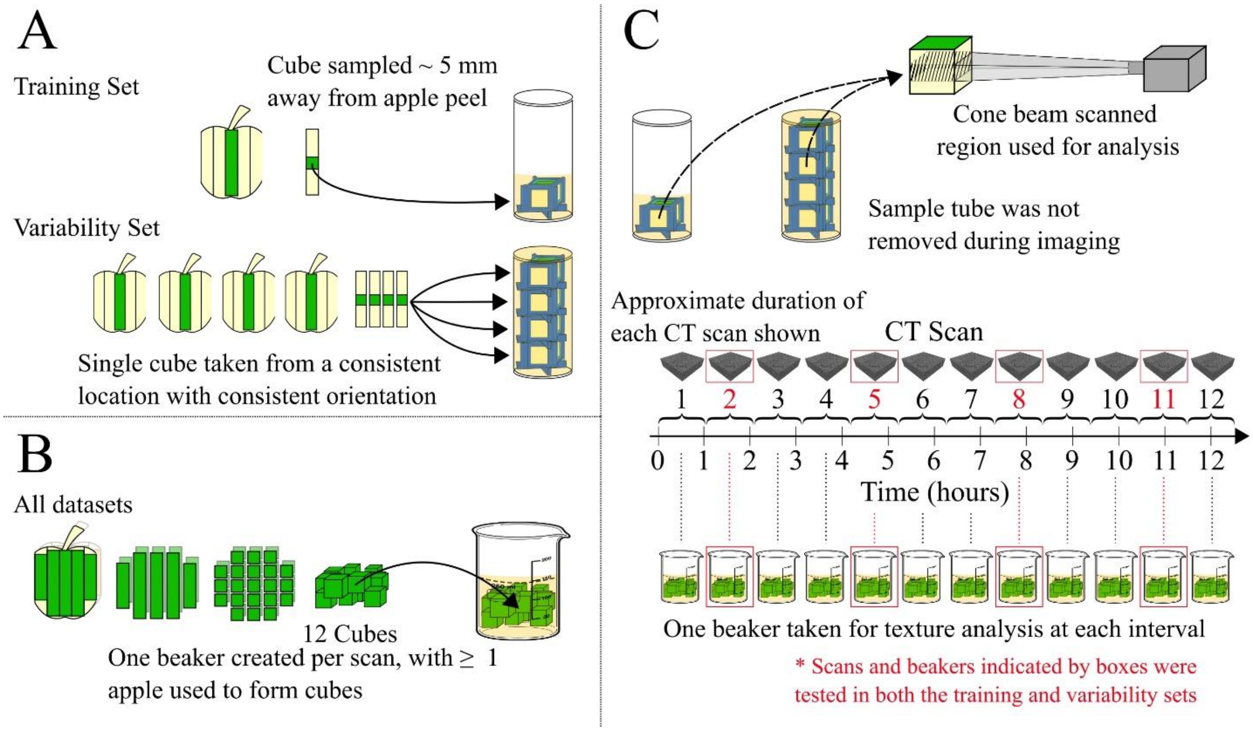

A
Training Set
Cube sampled ~ 5 mm away from apple peel
Variability Set
Single cube taken from a consistent location with consistent orientation
B
All datasets
12 Cubes
One beaker created per scan, with ≥ 1 apple used to form cubes
C
Cone beam scanned region used for analysis
Sample tube was not removed during imaging
Approximate duration of each CT scan shown
CT Scan
1
2
3
4
5
6
7
8
9
10
11
12
0
1
2
3
4
5
6
7
8
9
10
11
12
Time (hours)
One beaker taken for texture analysis at each interval
* Scans and beakers indicated by boxes were tested in both the training and variability sets

Figure 4.1. Collection of image and destructive data. A) Apples for both the training and variability sets were cut into cubes and placed in a cylindrical container for CT imaging. The side of the apple cube shown in green was placed such that it was facing vertically upwards in the sample holder for imaging. B) For collection of compression curves, apples were cut and placed into beakers for digestion in a temperature-controlled incubator for times up to 12 hours for both the training and variability sets. C) Micro-CT scans and concurrent digestion occurred over a span of approximately 12 hours, with either a single apple cube scanned repeatedly (training set) or multiple apple cubes scanned in series (variability set). The samples taken for texture analysis were selected at the mid-point time of each successive micro-CT scan.

4.3.1. Raw Materials

Data were derived from two sets of apples. The first set of apples was used to create a predictive model and is referred to as the training set. Raw materials and unprocessed data collected from the materials in the training set is described in Section 3.3.1. A second set of apples referred to as the variability set, which has not been previously described, was collected to validate the performance of the predictive model.

4.3.1.1. Training Set Raw Materials

Two volume bushels of size 56 apples (var. *Granny Smith*) were acquired from a local produce wholesaler. Apples were stored at 1 °C for up to five weeks. Simulated gastric juice was formulated according to Bornhorst and Singh, 2013.

4.3.1.2. Variability Set Raw Materials

Apples (var. *Granny Smith*) of varying sizes were acquired in small batches from a local produce vendor. Apples were stored at 1 °C for up to five weeks. Simulated gastric juice used for incubation was formulated as described Section 3.3.1.

4.3.2. 4.3.2. Data Collection

4.3.2.1. Sample Preparation and Data Collection for Training Set Data

Detailed sample preparation and data acquisition methods can be found Section 3.3. Briefly, a single cube of apple tissue was cut into a 12.7 mm cube and placed in a 3D-printed holder within a plastic tube (Fig. 4.1A). Immediately before image acquisition, 8.75 mL of either simulated gastric juice or deionized water was added to the sample tube.

Images were acquired with a Scanco uCT 35 Evaluation System (Scanco USA, Wayne, PA). Each 3D image required approximately 1.05 h to complete, after which the same region of the apple was scanned 11 more times for a total of 12 scans (Fig 4.1C). Reconstructed and cropped images represented a 11.1 × 11.1 × 4.3 mm section from the center of the apple cube in 18.5 um isotropic resolution (x y z = 601 × 601 × 232 px) and 16-bit depth. Scanning was performed in triplicate for each treatment for a total of 6 apples.

For collection of compression data, apples were subjected to similar conditions to those inside the CT instrument. A single beaker of 12 cubes was incubated at times representing the midpoint of each CT scan at 33 °C (Fig. 4.1C). After each incubation time, 10 of the 12 cubes in a given beaker were individually compressed at 1 mm/s to 50% strain. Incubations in simulated gastric fluid and water were each performed in triplicate.

In addition to the CT scans of apples submerged in liquid, scans of undigested apples were also recorded. These scans were not used for model training, but the apples were taken from the same batches used for the training set scans. Scans were performed in triplicate, using three unique apples. Compression data corresponding to these scans was collected from apples immediately after cutting.

4.3.2.2. Sample Preparation and Data Collection for Variability Set Data

Sample preparation in the variability set closely follows Section 3.3.1, with minor modifications to increase the biological variability present in the dataset.

For collection of image data, four cubes of apple tissue taken from four distinct apples were placed in an individual 3D printed holders within a sample tube, and the tube was filled with 35 mL of liquid (simulated gastric juice or deionized water), representing 8.75 mL per cube (Fig. 4.1A). Samples were scanned using the same imaging parameters as the training dataset (Fig. 4.1C). However, instead of 12 time-series scans from the center region of a single apple cube, the center sections of each of the four unique apple cubes were scanned at times corresponding with the 2nd, 5th, 8th, and 11th scans in the training set (Fig. 4.1C). Scans for each treatment were conducted in triplicate. Therefore, 24 apples in total were scanned in the variability set (3 replicates of 4 apples in 2 treatments). Triplicate scans of undigested apple cubes were also recorded from the variability set.

Compression profiles were collected from apples within the variability set at times corresponding with the midpoint of each of the four scans (1.60, 4.80, 8.00, and 11.20 h), using the same method as described in Section 4.3.2.1. Likewise, incubations were performed in triplicate for each simulated gastric juice and deionized water.

4.3.3. Data Processing

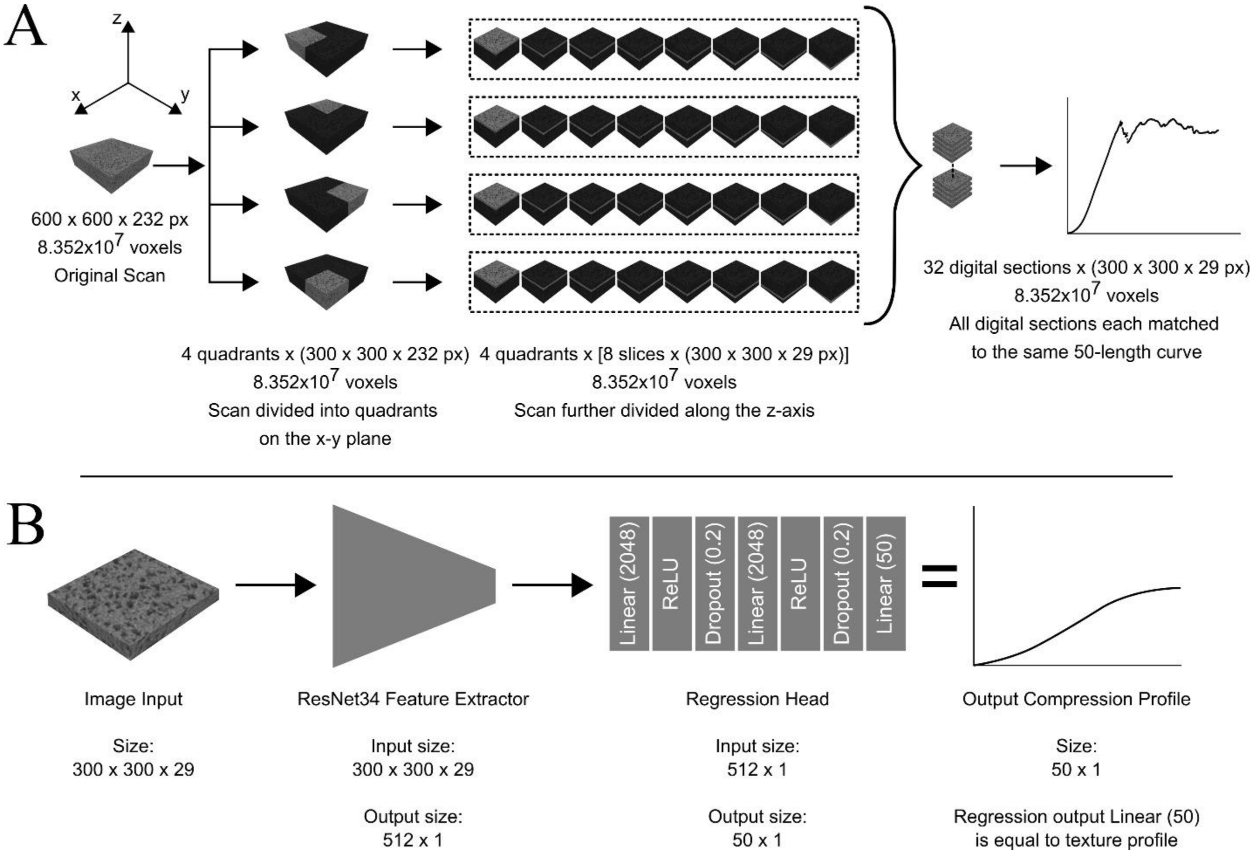

A
z
x
y
600 x 600 x 232 px
8.352x10^7 voxels
Original Scan
4 quadrants x (300 x 300 x 232 px)
8.352x10^7 voxels
Scan divided into quadrants
on the x-y plane
4 quadrants x [8 slices x (300 x 300 x 29 px)]
8.352x10^7 voxels
Scan further divided along the z-axis
32 digital sections x (300 x 300 x 29 px)
8.352x10^7 voxels
All digital sections each matched
to the same 50-length curve
B
Linear (2048)
ReLU
Dropout (0.2)
Linear (2048)
ReLU
Dropout (0.2)
Linear (50)
=
Image Input
Size:
300 x 300 x 29
ResNet34 Feature Extractor
Input size:
300 x 300 x 29
Output size:
512 x 1
Regression Head
Input size:
512 x 1
Output size:
50 x 1
Output Compression Profile
Size:
50 x 1
Regression output Linear (50)
is equal to texture profile

Figure 4.2. A) Sampling strategy used on each scan of an apple to augment image data and provide additional samples for model training. Each scan of apple tissue was divided into four quadrants in the x-y plane and 8 slices along the z-axis. These 32 resulting digital sections were associated with a single compression curve and treated as distinct samples for the purpose of model training. B) Model architecture. One digital section from an apple scan was used as the input image. A ResNet34 network modified to accept 29-channel images was used as a feature extractor, after which the features were sent to the regression head for prediction of a 50-length vector. This final vector represented the predicted compression curve. Dimensions below the labels represent the size of the data coming into and out of each step.

4.3.3.1. Pre-Processing

Before model training, the data were arranged such that a given scan of an apple cube corresponded with a single compression curve. This was done by matching each replicate of scanned apple tissue with its corresponding replicate of destructively tested apple cubes. Within a replicate, each 3D scan corresponded to approximately 10 compression curves from the same set of apples. These compression curves were converted to a single curve by taking the median force value at each distance up to 5.524 mm, representative of the minimum distance to which all compressions were performed. To minimize the size of the predicted curve while maintaining sufficient resolution, the curves were down-sampled from 1105-length (200 Hz) to 50-length vectors (9 Hz). After processing, the training set uniaxial compression data consisted of 72 compression curves, one corresponding to each of the 72 scans from 2 treatments, 12 time points, and 3 replicates.

Given the relatively small training dataset size in this study, data augmentation was used to increase training data variability and reduce the likelihood of overfitting (Simard et al., 2015; Yaeger et al., 1997). Augmentation was accomplished by digitally sectioning the 3D image from each scan, producing additional samples for training. Specifically, each 601 × 601 × 232 px 3D image was split into 32 sections representing 300 × 300 × 29 px digital sections with the assumption of axisymmetric tissue changes over the course of incubation (Fig 4.2A). This assumption was made based on the geometry of the experimental setup, in which the apple cube was in contact with the incubation liquid on all sides. Each of these 32 digital sections for each scan was associated with the same 50-length compression curve. Therefore, after data augmentation, the training dataset consisted of 2,304 digital sections, which were associated with 72 compression curves.

Data were handled the same way within the variability set as well as the undigested samples, where each scan was matched with a median compression curve. Data augmentation was also similarly performed. Therefore, the variability dataset consisted of 768 digital sections derived from 2 treatments, 4 scans, and 3 replicates, which were associated with 24 compression curves. The undigested samples from each of the training and variability sets consisted of 96 digital sections (32 digital sections per scan, 1 scan per replicate). These 96 digital sections per dataset were associated with compression curves collected from apples immediately after slicing. Undigested samples were not used to train the model, and association of image and compression data was done for performance assessment only.

4.3.3.2. Model Architecture and Training

A CNN was designed to accept 300 × 300 × 29 px input images and output a 50-length vector representing the compression curve (Fig. 4.2B). This was accomplished by first modifying the input layer of a ResNet34 model (He et al., 2016). Additionally, two linear layers were added in place of the classification layer, each followed by a rectified linear unit (ReLU) nonlinearity (Nair and Hinton, 2010) and a 20% dropout layer (Srivastava et al., 2014). The final 50-length vector was taken as the predicted compression curve, and loss values for training via backpropagation were calculated as the mean squared error between the final vector and the measured curve.

Models were trained using the PyTorch framework (Paszke et al., 2019). During training, input images were further augmented using random vertical and horizontal flips across the x-y plane, as well as a random number of 90-degree rotations.

4.3.4. Saliency Mapping for Interpretation of Model Results

Saliency mapping was performed using the technique of Gradient-weighted Class Activation Mapping (Grad-CAM) with some modifications to gain an understanding of the regions of the images the model used to make predictions (Selvaraju et al., 2016). This technique is typically used to generate a map representing regions of the input image with proportionally high attention given by the model for classification of the image into a specific class. However, in the current study, the entire output compression curve is meaningful. As a result, the value used for saliency mapping was selected to be the peak value of force in the curve. This value was also referred to as sample hardness (Drechsler and Bornhorst, 2018; Kong and Singh, 2008). As an additional change from the cited method, the ReLU transformation used to eliminate negative values in the original work was not used.

Visualization of the saliency maps was accomplished by generating a map for each of the 32 digital sections for a given scan. Output maps were rotated around the z-axis according to their original position and concatenated in the x-y plane. The 8 sections along the z-axis were averaged, producing a single saliency map for each scan, with pixel values comprised of average feature map weight x activation values, representing relative attention of the model in different regions of the input image (Selvaraju et al., 2016). Saliency maps were additionally produced for the variability set and arranged in the same fashion.

4.3.5. K-means clustering for image visualization

Following the procedure in Olenskyj et al. (2020), 3D scans were binarized using k-means clustering (Arthur and Vassilvitskii, 2007). Clustering was performed using the z-axis vector at each (x,y) location assigning each (x,y) pixel to one of two groups. This process was performed for visualization purposes and for comparison with saliency maps for the training and variability set images.

4.3.6. Image Analysis of Training and Variability Set Data

To quantify potential differences between the training set, variability set, and undigested apples, the intensity, porosity, and median pore diameter (d50) were calculated as described in Section 3.3.4.1 for each digital section within each dataset and compared across the datasets.

4.3.7. Performance Assessment

All listed performance values are represented by mean ± standard deviation unless otherwise noted. Model performance on the training set was assessed via 3-fold cross-validation across the three replicates (Guo et al., 2017) of both image and compression data, using root mean squared error (RMSE), mean average percent error (MAPE), and R^2. RMSE, MAPE, and R^2 were calculated between the median of all 50-point curves generated from each digital section within a scan and the median 50-point curve from all compressions at the treatment × time combination within a replicate.

For the training set data, performance on each replicate was determined after training with the two remaining replicates. From the two replicates used for training, data were randomly placed in training and validation sets using a 70/30 split. The held-out replicate was treated as the testing set for performance evaluation. This method of performance assessment resulted in the generation of three sets of model weights. Evaluation on the undigested samples and the variability set was performed using an ensemble approach, where predictions from all three sets of weights were averaged to obtain a single predicted compression curve for each sample. This was possible since no images of these apples were used in model training.

In addition to evaluation of the compression curves in their entirety, the hardness (peak force) value was extracted from each predicted curve and compared with the peak force of the measured compression curve. Hardness was extracted from the predicted compression curve from each of the 32 digital sections within each scan, which were considered subsamples for each treatment × time × replicate combination. Likewise, hardness was extracted from each measured compression curve, where the 10 compressions performed for each time within each replicate were treated as subsamples. This comparison between measured and predicted data was analyzed by fitting the curve to a modified Weibull model (Eqn. 3.2). However, unlike in Section 3.3.5.3, in addition to the k (h^{-1}) and β (dimensionless) values, H_0 (N) was fit as well.

To improve model fits, hardness from the first scan was adjusted to 0 h to become the initial hardness and other values of time were translated accordingly by subtracting the first x-value (0.53 h for the training and 1.60 h for the variability set) from all x-values. Therefore, in Eqn. 3.2, H_0 is an estimate of the hardness of the first measured time point.

From the fit curves, 95% prediction intervals were calculated using *nlpredci* in MATLAB for both measured and predicted curves. Prediction intervals from measured and predicted data were used to calculate the intersection over union (IOU) between the prediction intervals at each time point, where 1 represents perfect accordance and 0 represents non-overlapping intervals. Hardness values were also extracted from the variability set images and compared with their destructive values. Subsamples within both the image and measured data of the variability set were handled the same way as in the training set.

4.4. Results and Discussion

4.4.1. Model Performance within the Training Set (Cross-Validation)

4.4.1.1.Compression Curve Prediction

Model performance at each time point is shown in Figure 4.3. Notably, the model can accurately predict both the sigmoidal shapes of the water and gastric time points at < 4.80 h, as well as the more complex curves with multiple points of inflection, such as the 11.20 h curve in the first replicate of the gastric treatment. The capability of the model to express various curve shapes demonstrates one of the advantages of predicting the entire curve, rather than fit parameters or scalar values of force.

Overall RMSE of cross validation within the training set was 4.36 N (5.04 and 3.55 N for gastric and water samples, respectively). Overall MAPE of cross validation across all treatments was 26.4 % (39.7 and 13.0 % for gastric and water samples, respectively). The R^2 calculated on all points in all curves was 0.939 (0.919 and 0.956 for gastric and water samples, respectively).

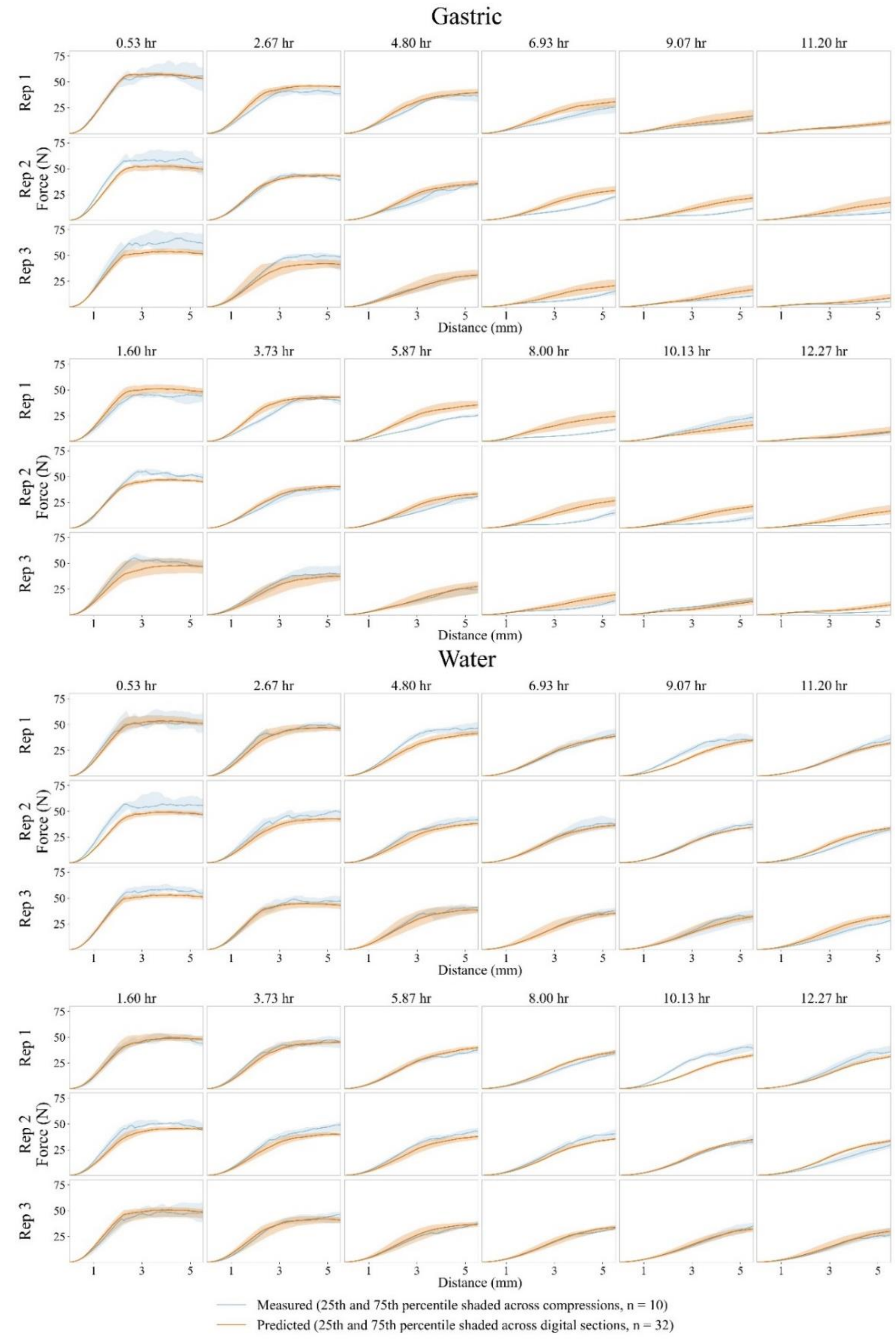


Figure 4.3. Measured compression curves and corresponding predicted curves. All curves represent compression to 5.524 mm via a 50-point down-sampled vector. Each row within gastric and water treatments represents a replicate (3 replicates/treatment). Predicted curves for a given replicate were generated using a model trained on data from the other two replicates (cross-validation). Curves represent the median, with the shaded areas representing the 25th and 75th percentile of 32 digital sections or 10 compressions for predicted and measured curves, respectively.

The consistency of RMSE across time is expected (Fig. 4.4), due to the mean squared error objective function used to train the models. Specifically, the model was trained to minimize squared error across all samples, which were treated independently. However, MAPE was magnified in gastric samples at later time points due to the low values of force (Fig. 4.3). Specifically, within the first four time points ($t \leq 4.8$ h) MAPE of the gastric samples was 13.2 % on average, while MAPE of the water samples at the same time points was similar at 10.9 %. However, within the last four time points ($t \geq 9.07$ h) MAPE of the gastric samples rose to 56.5 %, while the water samples only increased to a MAPE of 17.8 %.

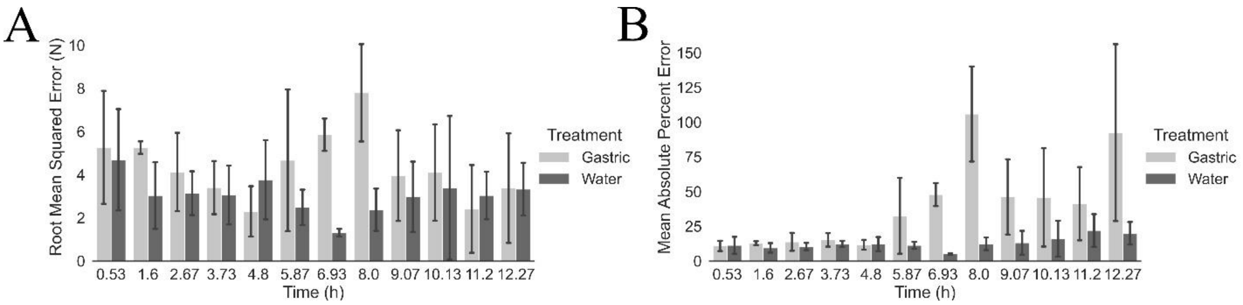


Figure 4.4. RMSE (A) and MAPE (B) values for predicted vs. measured compression curves. Each RMSE and MAPE value represents the total error between a ground truth 50-length vector represented by the median of all compressions within a time point of a replicate and a predicted vector represented by the median of predictions from all 32 digital sections of a scan within a replicate. Predicted curves for each replicate were generated using a model trained on the two other replicates. Error bars represent standard deviation between replicates (n = 3).

The effect of force magnitude on performance can be further localized by the level of displacement within the curves (Table 4.1). At the highest level of displacement (4.5 – 5.5 mm), force in the water treatment demonstrated a 34.1 N difference between the lowest and highest measured values. On the other hand, gastric treatments demonstrated a range of 64.3 N. The range of force in the gastric treatment was consistently greater than the water treatment at all levels of displacement, indicating greater variation between measured force during incubation in simulated gastric juice. This increased range may have led to the systematic overprediction of force values in gastric time points > 5.87 h, as well as underprediction in both treatments in time points < 4.80 h (Fig. 4.3). Similarly, the consistently low values of force at low levels of displacement likely led to lower error compared with the later segments of the curves, due to the small variation between samples in the first part of each compression curve (Table 4.1). Within a displacement < 1.1 mm, the measured force used for model training showed a range of 20.1 N across all gastric samples and 19.0 over all water samples ($n = 36$ each). Within all curve segments after the first, the range of force values was larger (34.1 to 64.9 N across both treatments).

In a work similar to the current study, Liang et al. (2017) predicted stress-strain vectors from microscopy images of collagenous tissue. As opposed to the end-to-end prediction used in this work, the authors first parameterized measured curves using PCA to obtain a compact representation, then trained a CNN to predict three fit parameters. The approach allowed for comparatively low prediction error (10.9 - 12.6%), which is similar to the error in the water treatment and early gastric time points ($t \leq 4.8$ h) in this study (13.0 and 13.2 %, respectively). However, the results are not directly comparable due primarily to differences in study design. Liang et al. (2017) used 48 independent samples imaged on a smaller length scale using confocal laser scanning microscopy. The stress-strain curves used for prediction were consistent in shape, and therefore they were more easily parameterized. Additionally, the imaged samples were identical to the samples used to generate ground truth mechanical testing data. On the other hand, the present study included time-series data on fewer biological samples (6), but a greater number of ground truth compression curves (72). Furthermore, ground truth data from this study was not collected from the same samples as the images. Additional differences include variations in model development (unsupervised learning with supervised regression as opposed to fully supervised end-to-end prediction in this work), as well as the evaluation method (leave-one-out cross validation, as opposed to the 3-fold cross validation in this work).

Table 4.1. MAPE and RMSE of cross validation on different regions of the curve. Predicted 50-point curves were split into five segments of 10 points each, and error was calculated within each segment. Values represent comparisons between destructive and image curves, calculated by taking the median value at each level of displacement over all digital sections and compressions, respectively, of all scans within each replicate. Performance in MAPE and RMSE is represented by mean ± standard deviation of all 12 time points across all 3 replicates (n = 36).

Treatment	Displacement (mm)	Avg. Measured Force (N)	Measured Range (N)	MAPE (%)	RMSE (N)
Gastric	0 – 1.1	2.10 ± 1.85	20.1	17.0 ± 11.9	0.438 ± 0.457
	1.2 – 2.2	11.6 ± 11.6	53.7	25.8 ± 22.0	2.31 ± 1.96
	2.3 – 3.3	19.9 ± 18.7	60.7	50.6 ± 55.5	4.76 ± 3.02
	3.4 – 4.4	23.9 ± 19.2	64.9	58.3 ± 76.5	5.34 ± 3.8
	4.5 – 5.5	26.3 ± 17.8	64.3	47.0 ± 66.8	5.43 ± 3.84
Water	0 – 1.1	2.42 ± 1.75	19.0	16.9 ± 9.65	0.555 ± 0.612
	1.2 – 2.2	14.7 ± 10.3	51.9	15.4 ± 12.3	2.22 ± 2.17
	2.3 – 3.3	27.7 ± 13.5	50.0	13.0 ± 12.2	3.28 ± 2.75
	3.4 – 4.4	35.3 ± 11.0	43.4	10.6 ± 8.70	3.55 ± 2.40
	4.5 – 5.5	39.4 ± 8.36	34.1	9.02 ± 5.40	3.57 ± 1.97

In an additional study on prediction of non-scalar mechanical properties from images using deep learning, Herriot and Spear (2020) used a 3D CNN architecture to predict 2D maps of effective yield strength in simulated heterogeneous samples of metals. The 3D CNN demonstrated high prediction performance on a holdout set (R^2 of 0.95). However, to achieve this result, the model inputs were coupled with input volume auxiliary information. Still, the performance compares well to the R^2 values of cross validation obtained in this work, equal to 0.919 and 0.956 for gastric and water samples, respectively.

4.4.1.2. Comparison of measured and predicted hardness from compression curves

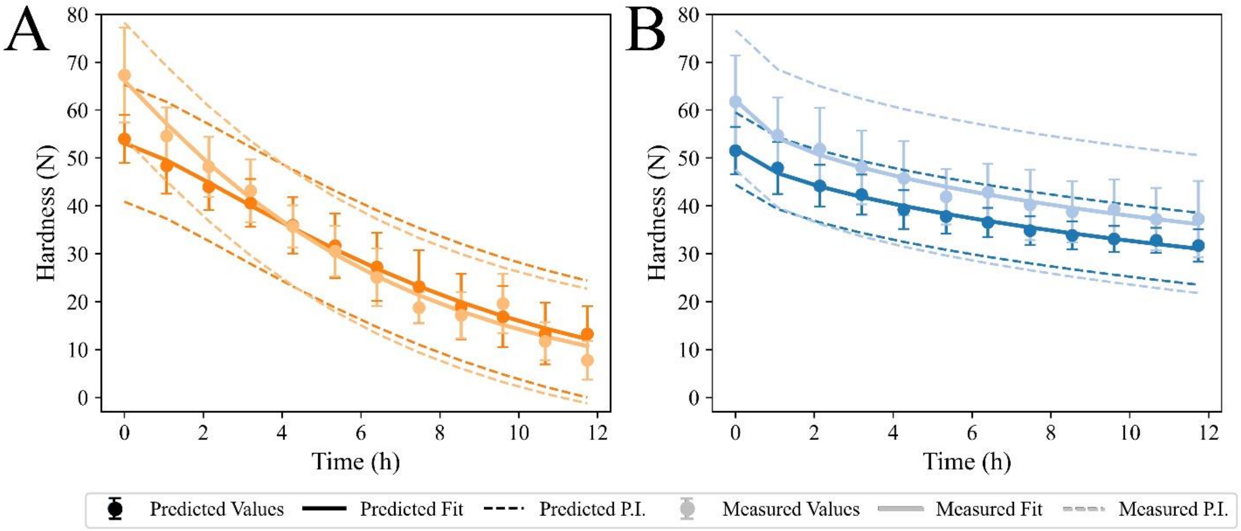


Figure 4.5. Hardness of gastric (A) and water (B) treatments predicted from image data and measured by destructive testing. Hardness from image data is averaged across all 32 digital sections within each scan of a replicate. Hardness from destructive testing is averaged over all compressions within each time point in a replicate. Data points represent the mean and error bars represent standard deviation between replicates for both image and destructive data (n = 96 and n=30, respectively). Solid lines represent model fits to Eqn. 3.2. Dotted lines represent 95% prediction intervals. Measured values are shown with lighter shading compared to predicted values that are shown with darker shading for both gastric and water treatments.

Predicted and measured hardness values are shown along with Weibull model curve fits (Eqn. 3.2) in Fig. 4.5. Although the fit curves were not identical between treatments (Table 4.2), overlapping prediction intervals at each time point in both gastric and water treatments demonstrate the similarities between predicted and measured hardness values. Specifically, IOU for the gastric and water treatments were 0.878 ± 0.146 and 0.512 ± 0.031, respectively (mean ± SD, n = 12).

Table 4.2. Fit parameters for measured and predicted values in gastric and water treatments in the training set. All values represent mean ± SEM of fit parameters for curves fit to all points within a treatment × dataset for each measured and predicted sets.

		Gastric		**Water**	
		Measured	*Predicted*	*Measured*	*Predicted*
Training Set	**H_0 (N)**	66.1 ± 1.03	53.1 ± 0.539	62.1 ± 1.31	52.0 ± 0.378
	k (h^{-1})	0.151 ± 3.37×10^{-3}	0.115 ± 1.50×10^{-3}	2.89x10^{-2} ± 4.03×10^{-3}	3.16x10^{-2} ± 1.47×10^{-3}
	β (-)	1.05 ± 4.05×10^{-2}	1.27 ± 4.02×10^{-2}	0.570 ± 6.12×10^{-2}	0.668 ± 2.67×10^{-2}
	R^2	0.889	0.823	0.490	0.719
Variability Set	**H_0 (N)**	54.6 ± 1.38	50.0 ± 1.19	53.8 ± 1.10	55.3 ± 0.496
	k (h^{-1})	0.179 ± 8.94×10^{-3}	3.01×10^{-2} ± 1.08×10^{-3}	2.69x10^{-2} ± 9.49×10^{-3}	3.34x10^{-2} ± 4.02×10^{-3}
	β (-)	1.09 ± 0.113	0.758 ± 0.184	0.581 ± 0.122	0.693 ± 5.84×10^{-2}
	R^2	0.869	0.207	0.655	0.707

The RMSE for predicted hardness over all time points was 6.41 N for gastric and 6.89 N for water samples. MAPE was 23.0 and 13.4 % for gastric and water samples, respectively. Similar to the full curves, within the gastric treatment, early time points (t < 4.8 h) were underpredicted, whereas later time points tended to be overpredicted. In contrast, the water samples were systematically underpredicted. The reason for this is unclear, although since gastric and water samples were both included equally in training, the lower hardness values in the gastric samples may have negatively influenced the predicted hardness of the water samples.

Compared to previous works in prediction of scalar bulk material properties from images of materials, the error values in the current study are higher, with previous studies demonstrating below 5% error (Li et al., 2019; Yang et al., 2018; Ye et al., 2019). However, previous studies explored simpler datasets. Specifically, input images were constrained to contain either two material classes with consistent, known material properties derived from simulated data (Yang et al., 2018; Ye et al., 2019) or five material classes derived from a parameterization of measured samples (Li et al., 2019). Contrarily, the present study consisted of unmodified images of biological tissue. Additionally, the model in this work was trained to predict the entire compression curve, potentially limiting scalar prediction performance. Finally, although time series analysis and data augmentation increased the number of training samples, the biological variability in the present study was limited to four unique apples (two per treatment) used for each fold of cross-validation. After training on only four apples, the model predicted compression curves for two unseen apples. In comparison, the previous studies mentioned above used between 5,000 and 20,000 synthetically generated samples in training to achieve low error values. Due to the difficulty involved in sample preparation for food analysis, the performance of the model in the current study shows potential for application in a wide range of foods and for processes where limited data may be available.

4.4.2. Saliency mapping with Grad-CAM

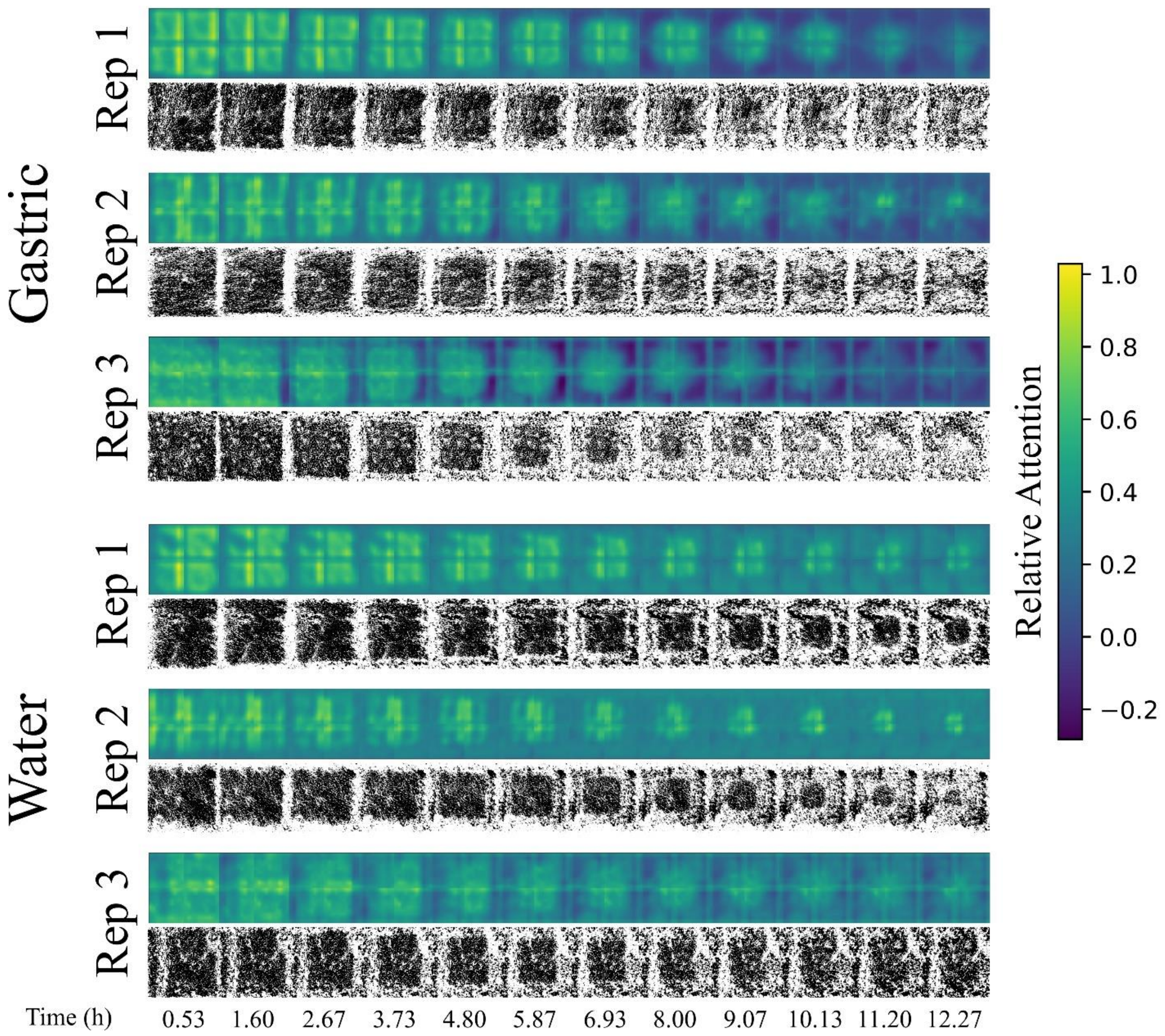


Figure 4.6. Saliency maps produced with Grad-CAM. Each row within a treatment represents a replicate incubation. Binary maps positioned below each set of saliency maps represent a visualization of pixelwise differences between z-axis vectors in the x-y plane (Olenskyj et al., 2020). Pixels within each saliency map represent weight × activation values (relative attention). Saliency maps are derived from all 32 digital sections of a scan within a replicate, where sections are averaged along the z-axis. The color bar (right) applies to all averaged maps within the training set.

Saliency maps produced with Grad-CAM (Fig. 4.6) showed similar trends across all replicates and treatments. The relative attention of the model with respect to prediction of a higher maximum force value was evenly distributed across the x-y plane of the input image for the early scans (e.g. 0.53, 1.60, and 2.67 hours). With increasing incubation time, saliency maps demonstrated more prominent attention in the center of the apple cube, with less dependency on the outer edges of the cube for prediction of a higher maximum force value. This trend suggests that the more important region of the image for predicting hardness is found closer to the center of the apple cube as time increases, which can be explained based on the diffusion-limited nature of digestion (Kong and Singh, 2009b; Olenskyj et al., 2020). Specifically, the material on the outer edges of the sample interacts with the incubation medium before the inner regions of tissue. Since this interaction results in a softening of the food matrix, as incubation time increases, the exterior layers of the sample contribute less to increasing the maximum force.

Visualization of this same dataset using k-means clustering as a method of binarization across the x-y plane shows a similar trend to the saliency maps using only intensity-based information from the micro-CT scans (Fig. 4.6). This relationship between image intensity and tissue hardness was demonstrated previously at the entire cube scale (Olenskyj et al., 2020). However, this previously established relationship was not specifically presented to the CNN. Instead, the end-to-end approach used to train the model caused this pattern of attention to emerge, suggesting that the model was able to converge to a mechanistic representation of the system from which physically meaningful parameters can be extracted.

4.4.3. Performance on out-of-distribution data

4.4.3.1. Characterization of Images from Different Datasets

Comparison of intensity and morphology parameters across the training and variability datasets for the gastric and water treatments did not reveal many systematic differences between the datasets (Table 4.3). However, the undigested samples from both the training and variability set demonstrated considerably lower median intensity values (2728 ± 172 and 2775 ± 227, respectively) when compared with the first scan of the training set (2871 ± 191). Median pore diameter was also substantially higher in the undigested samples (180.9 ± 34.6 and 170.3 ± 27.5 µm in the training and variability sets, respectively) as compared with the training set (89.24 ± 13.19 µm). This discrepancy may have been due to rapid water absorption by the food matrix into large pore spaces. Alternatively, the incubation medium (air vs. liquid) has been shown to significantly influence attenuation values of calibration phantoms, which may account for differences between undigested samples and samples in liquid (Nazarian et al., 2008).

Table 4.3. Comparison of intensity and morphology measures from micro-CT images from different datasets. Values represent Median ± IQR (75th – 25th percentile value).

Metric	**Training Set (Scans 2, 5, 7, 11)**	**Variability Set**
Gastric (n = 384 in both sets)		
Intensity (CT Value)	3062 ± 166	3010 ± 219
Porosity (Percent)	16.92 ± 3.38	17.01 ± 4.40
Median Pore Diameter (µm)	88.72 ± 7.95	82.44 ± 22.1
Water (n = 384 in both sets)		
Intensity (CT Value)	2986 ± 195	3011 ± 233
Porosity (Percent)	17.53 ± 4.59	16.10 ± 5.70
Median Pore Diameter (µm)	72.53 ± 9.58	78.86 ± 15.7

Metric	**Training Set (Scan 1; n = 192)**	**Undigested Samples from Training Set (n = 96)**	**Undigested Samples from Variability Set (n = 96)**
Intensity (CT Value)	2871 ± 191	2728 ± 172	2775 ± 227
Porosity (Percent)	20.15 ± 4.62	21.77 ± 5.12	20.25 ± 6.04
Median Pore Diameter (µm)	89.24 ± 13.19	180.9 ± 34.6	170.3 ± 27.5

Although few systematic differences between datasets were observed in the selected image features, current research efforts in understanding the features used by CNNs is ongoing (Barredo Arrieta et al., 2020), and future improvements in dissecting these models should allow for a more detailed analysis of the relationship between specific image features and performance.

4.4.3.2. Prediction on samples incubated in simulated gastric juice or water from the variability set

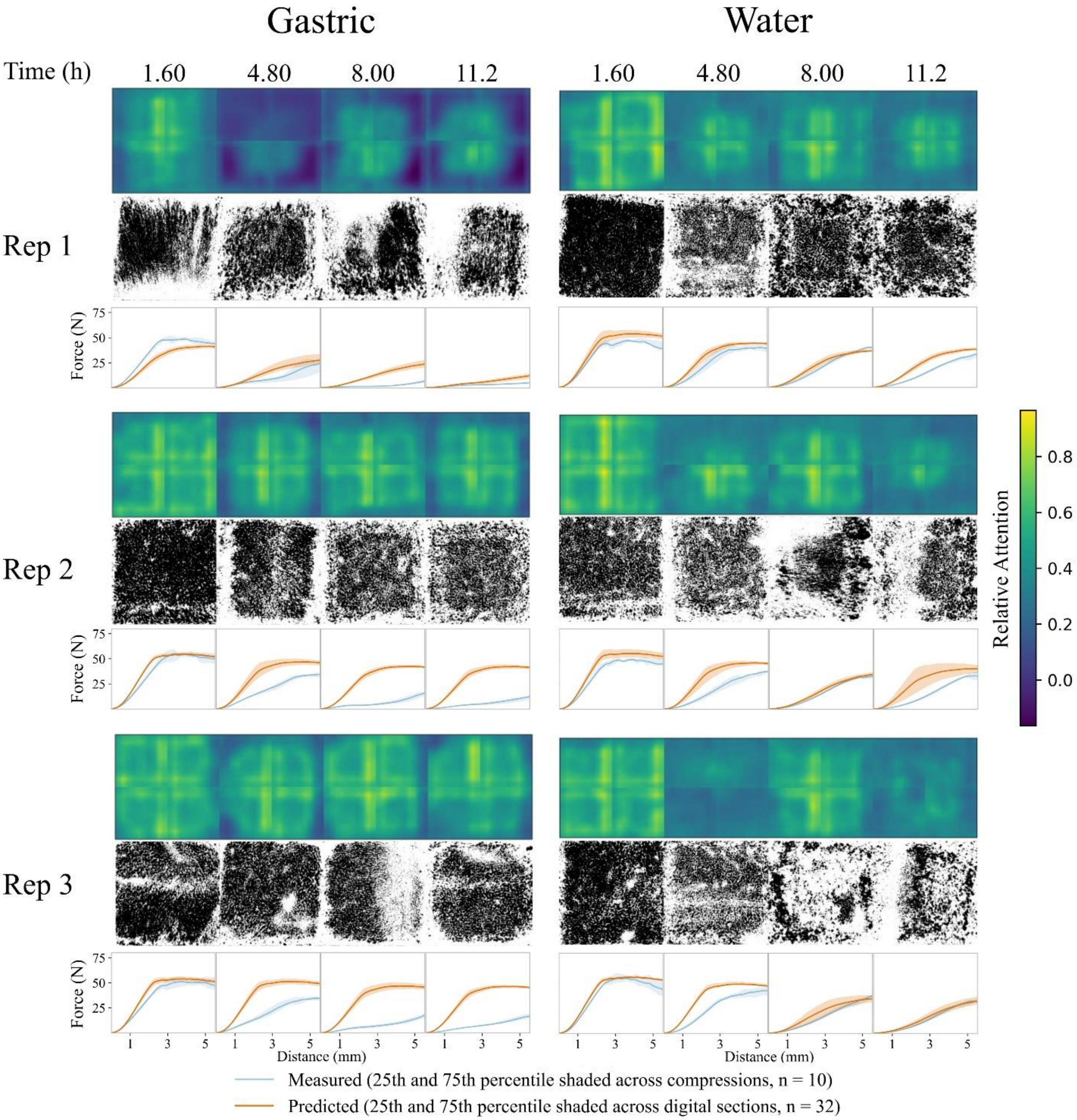


Figure 4.7. For each replicate of each treatment, the first row represents saliency maps produced with Grad-CAM. Individual maps are derived from all 32 digital sections of a scan within a replicate and averaged across all three models ensembled to make predictions on the variability set. Sections are averaged along the z-axis. Pixels within each map represent weight × activation values (relative attention). The colorbar (right) applies to all averaged maps within the variability set. Binary maps shown in the row below each set of saliency maps represent a visualization of pixelwise differences between z-axis vectors in the x-y plane (Olenskyj et al., 2020). Measured compression curves and corresponding predicted curves are shown in the row below the binary maps for each

treatment. All curves represent compression to 5.524 mm via a 50-point down-sampled vector. Predicted curves for a given replicate were generated by ensembling predictions from the three sets of model weights used for the training set. Curves represent the median of 32 digital sections or 10 compressions for predicted and measured curves, respectively, with shaded regions representing the 25th and 75th percentile values. Measured values are shown with lighter shading compared to predicted values that are shown with darker shading for both gastric and water treatments.

The RMSE for prediction on the variability set was 14.3 N across all replicates, treatments, and times (n = 24). MAPE was 129% and R^2 was 0.296 across all replicates.

Within the variability set (Fig. 4.7), performance in the water treatment was improved over the gastric treatment. RMSE was 18.5 N in the gastric treatment and only 8.14 N in the water treatment. MAPE was 214 % and 44.4 % in gastric and water treatments, respectively (n = 12 for comparisons within a treatment). Performance broken down across time points is displayed in Fig. 4.8. Notably, predicted curves appeared similar between treatments, with predicted gastric curves appearing similar to measured water curves at each time point. This similarity between treatments suggests that the out-of-distribution tissue structure in the variability set reduced the ability of the model to discriminate between tissue changes due to gastric juice or water, even though limited differences were observed in the intensity and morphology parameters between the training and variability set images (Table 4.3).

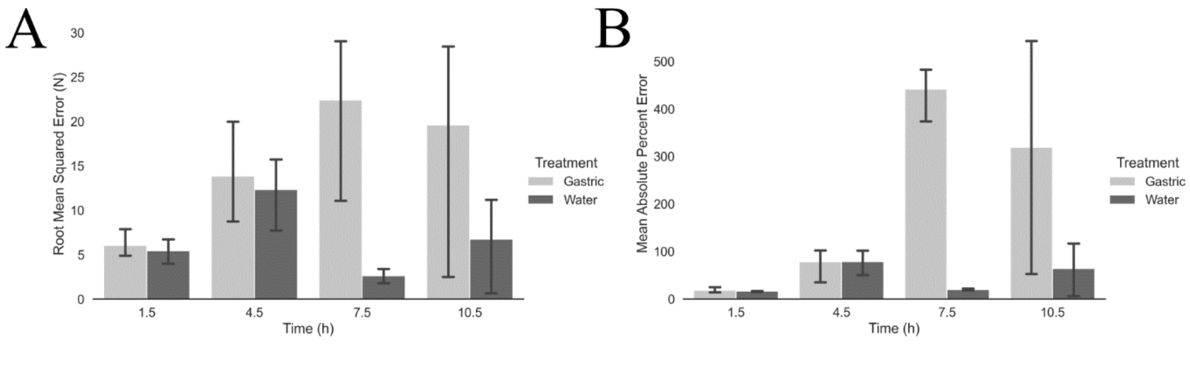


Figure 4.8. RMSE (A) and MAPE (B) values for predicted vs. ground truth texture curves within the variability set. Each RMSE and MAPE value represents the total error between a ground truth 50-length vector calculated based on the median of all compressions within a time point (scan) of a replicate and a predicted vector calculated based on the median of all predictions from all 32 digital sections of a time point (scan) within a replicate. Predicted curves for each replicate were generated using the average of curves predicted from all three models trained on the training set data. Error bars represent differences between replicates (95% CI, n = 3).

The overall error in the variability set was driven primarily by poor predictive performance on gastric samples. This performance drop was expected, since the variability set was composed of smaller batches of apples from a different source compared to the training set. Saliency maps with Grad-CAM provided insight into attention of the network during prediction (Fig. 4.7). As demonstrated by the saliency maps, model attention did not shift inwards for the later incubation times, in contrast to the observed trends in the training set (Fig. 4.6). Attention in the gastric treatment was consistently even across the input images in replicates 2 and 3. However, model attention did shift inwards over time for the first replicate in the gastric treatment. This shift in attention is likely responsible for the improved performance in replicate 1 of the gastric samples (Fig. 4.7).

On the other hand, performance on the water samples was considerably higher, although error was still increased relative to the training set. However, the water samples within the training set contained only three apples, while the water samples in the variability set came from 12 apples from a different commercial source and over a larger potential range of harvest and storage conditions. Within that context, performance on the water samples in the variability set demonstrated the robustness of the CNN approach, which would likely be improved with an increased amount of training data.

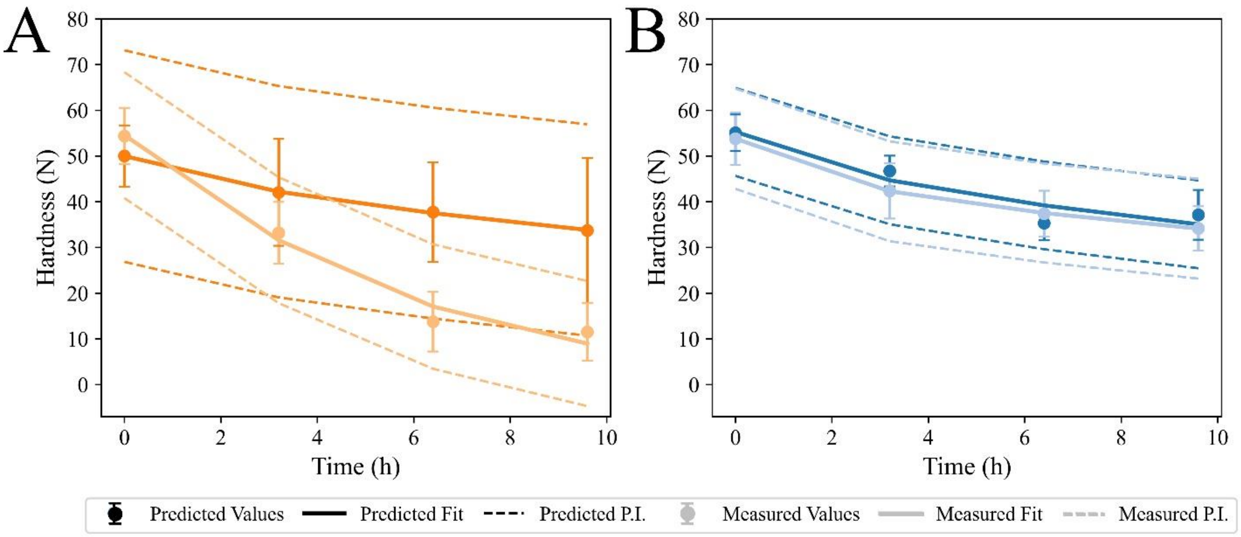


Figure 4.9. Hardness of gastric (A) and water (B) treatments predicted from image data and measured by destructive testing. Hardness from image data is averaged across all 32 digital sections within each time point of a replicate. Predictions from each digital section are derived from ensembling all three sets of model weights trained on the training set. Hardness from destructive testing is averaged over all compressions within each scan in a replicate. Error bars represent standard deviation between replicates for both image and destructive data (n = 96 and n=30, respectively). Solid lines represent model fits to Eqn. 3.2. Dotted lines represent 95% prediction intervals. Measured values are shown with lighter shading compared to predicted values that are shown with d arker shading for both gastric and water treatments.

Hardness values extracted from the predicted compression curves in the variability set further demonstrated decreased predictive performance of the model (Table 4.2) on the gastric treatment of the variability set (Fig. 4.9). The wide prediction interval in the gastric samples, suggests that the model was unable to map the out-of-distribution input to the correct value of hardness. IOU for the gastric and water treatments was 0.407 ± 0.171 and 0.846 ± 0.034, respectively (mean ± SD, n = 4). Notably, while the IOU for the gastric samples within the variability set was reduced as compared with the training set (0.407 vs. 0.846, respectively), the IOU for the water treatment was higher in the variability set compared to the test set (0.512 vs. 0.846, respectively).

The overall RMSE for predicted hardness over all time points was 18.54 N for gastric and 4.04 N for water samples. MAPE was 102.4 and 8.57 % for gastric and water samples, respectively. In contrast to the gastric samples, water samples in the variability set demonstrated a higher performance in hardness prediction, compared to the training set. Unlike studies which focus on cross-validation metrics or performance on a holdout set from the same data source, this study makes use of two distinct sets of materials. For comparison, Li et al. (Li et al., 2019) used synthetically generated mesoscale images for training and real images for testing, while predicting effective modulus of shale samples. Percent error in their study increased from 0.55 to 0.97 % between cross-validation and test sets, echoing the performance decrease seen in this study in the gastric samples.

Future development of this CNN-based approach with a wider variety of model foods is necessary to elucidate the specific attributes in the input images, which may impact model performance.

4.4.3.3. Curve Prediction in Undigested Samples

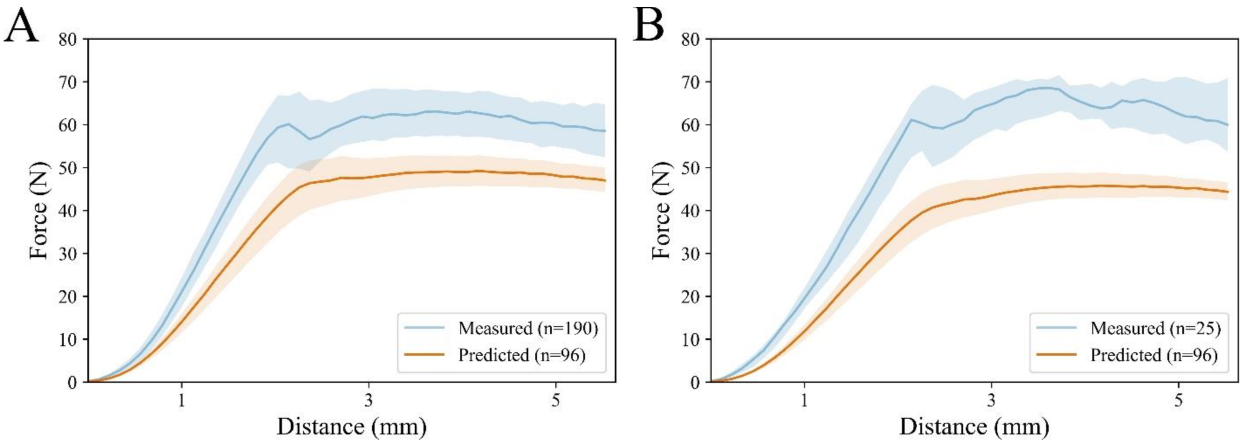


Figure 4.10. Measured and predicted compression curves from undigested samples in the training (A) and variability (B) sets. Predictions were generated via an ensemble approach by predicting the curve for each digital section using all three sets of model weights from the training set, then averaging the three curves to create a single 50-point vector for each of the 96 digital sections from images from undigested apples. The central curve for both predicted and measured sets was generated by taking the median of all digital sections and compression curves, respectively. Shaded regions represent the 25th and 75th percentile within predicted and measured curves. Measured values are shown with lighter shading compared to predicted values that are shown with darker shadin g.

The overall RMSE for prediction on undigested samples in the training set was 12.10 N between the averaged predicted curve after ensembling and the average of the destructive curves collected on undigested apple cubes. This value was considerably higher than the performance under cross-validation (4.36 N). R^2 on the undigested apples from the training set was 0.683, which represents a performance decrease from an R^2 of 0.939 in the corresponding incubated apples. Contrarily, MAPE on the undigested training set samples was 25.8 %, which is similar to the value found under cross-validation (26.4 %). The RMSE for the undigested apples from the variability set was 16.98 N, MAPE was 35.0 %, and R^2 was 0.449.

Like the trends seen in the incubated samples from the training set, underprediction in the undigested samples may have been driven by the high hardness in these samples. Additionally, within the undigested samples, the lack of incubation liquid present while imaging likely affected the intensity values of the images due to reduced attenuation of the X-ray beam, which influenced the image properties (Table 4.3). Difference in mean intensity may be avoided by normalizing input images, which is common practice in training deep learning models (Sudeep and Pal, 2017). However, preliminary trials suggested that scaling input intensity values negatively impacted model performance under cross-validation. In this case, due to the relationship between CT intensity value and density (Baker et al., 2012), maintaining raw CT values likely allowed the model to use overall intensity differences between samples to guide predictions.

Despite differences in acquisition method, visual comparison of the compression curves (Fig. 4.10) demonstrates that the micro-CT imaging and CNN prediction model was able to yield a similar curve shape to the measured hardness data. This result further suggests promise for future application of this approach to characterize food mechanical properties across different foods, during processing or digestion processes.

4.5. Conclusions

The combination of nondestructive micro-CT imaging with rapid prediction via CNNs can be employed to analyze time series data, such as data collected during in vitro gastric digestion, with minimal raw materials and experiment time. Within a model incubation system representing in vitro gastric digestion of apples, a CNN model designed to predict compression curves from 3D image inputs demonstrated high performance under cross-validation. The relationship between predicted and measured uniaxial compression curves had an R^2 value of 0.939 and RMSE of 4.36 N. Outside of the cross-validation dataset, predictive performance declined, with an R^2 of 0.296 and RMSE of 14.3 N. Nevertheless, high predictive performance was maintained on compression curves from water samples. Analysis of model attention using saliency maps within the cross-validation data demonstrated that model attention aligned with expectations based on mechanistic properties of the dynamic system, wherein structural breakdown progressed from the outer faces of the sample inwards. This trend was not observed on out-of-distribution data in the gastric samples, which aligned with the reduced model performance. Overall, the direct prediction of uniaxial compression curves from micro-CT image data using CNNs is a promising technique which may be applicable to a variety of time-series food processes. Future research is necessary to validate this approach in more food systems and food processing steps, as well as to improve the robustness of the approach on new data.

CHAPTER 5. END-TO-END DEEP LEARNING FOR DIRECTLY ESTIMATING GRAPE YIELD FROM GROUND-BASED IMAGERY

5.1. Abstract

Yield estimation prior to harvest is a powerful tool in vineyard management, as it allows growers to fine-tune management practices to optimize yield and quality. However, yield estimation is currently performed using manual sampling, which is time-consuming and imprecise. This study demonstrates the applicability of nondestructive proximal imaging combined with deep learning for yield estimation in vineyards. Continuous image data collection using a vehicle-mounted sensing kit combined with collection of ground truth yield data at harvest using a commercial yield monitor allowed for the generation of a large dataset of 23,581 yield points and 164,699 images. Moreover, this study was conducted in a commercial vineyard which was mechanically managed, representing a challenging environment for image analysis but a common set of conditions in the California Central Valley. Three model architectures were tested: object detection, CNN regression, and transformer models. The object detection model was trained on hand-labeled images to localize grape bunches, and detections were either counted or their pixel count was summed to obtain a metric which was correlated to grape yield. Conversely, regression models were trained end-to-end to directly predict grape yield from image data without the need for hand labeling. Results demonstrated that both a transformer model as well as the object detection model with pixel area processing performed comparably, with a mean absolute percent error of 18% and 18.5%, respectively on a representative holdout dataset. Saliency mapping was used to demonstrate the attention of the CNN regression model was localized near the predicted location of grape bunches, as well as on the top of the grapevine canopy. Overall, the study demonstrated the applicability of proximal imaging and deep learning

for prediction of grapevine yield on a large scale. Additionally, the end-to-end modeling approach was able to perform comparably to the object detection approach while eliminating the need for hand-labeling.

5.2. Introduction

Yield estimation is a valuable tool for crop management and precision agriculture. Estimation of crop yield in advance of harvest allows growers to make more informed decisions in negotiating pricing contracts or allocating quantities of crop for sale, and may also inform crop management decisions (Diago et al., 2012; Nuske et al., 2014). There are many approaches which can be taken towards estimating crop yield, the most common method involving manually sampling from a very small percentage (e.g., approximately 1%) of plants and extrapolating the yield data to the entire field. While this process is comparatively low-cost, it can be both time-consuming and imprecise (Liu et al., 2020; Wulfsohn et al., 2012). As opposed to manual sampling, remote sensing can capture information from an entire field in a short period of time, whether via satellite or via unmanned aerial vehicles (UAVs). The advantages of satellite-based remote sensing are speed, coverage, and a lack of requirement for manual labor. However, the resolution of publicly available satellite-based remote sensing is poor, typically between 10 and 30 m2 (Khaliq et al., 2019). Atmospheric conditions can also be an issue as clouds, smoke, and haze can obscure the area of interest. Additionally, data collection cannot be scheduled, due to the regular orbit of the satellite. Remote sensing images acquired using UAVs can provide increased resolution, but requires a trained pilot (Khaliq et al., 2019; Yang et al., 2019). Additionally, for specialty crops where fruit is located beneath vegetative cover, such as tomatoes and grapes, the overhead angle of images creates difficulties in imaging, especially when foliage is dense (Di Gennaro et al., 2019).

Proximal imaging, referred to here as imaging from ground-based equipment, has also been extensively explored for yield estimation (Gongal et al., 2015). Proximal sensing using imagery can be advantageous, as high-resolution images from a lateral perspective can be collected from underneath a canopy. However, like UAV imaging, proximal imaging requires specialized equipment. Additionally, while the resolution and perspective of proximal imagery can allow for direct visualization and quantification of fruit, occluded fruits can lead to inaccuracy in the estimation (Dunn and Martin, 2004; Mu et al., 2020; Wang et al., 2021). Nevertheless, there has been extensive work in image-based proximal sensing for fruit detection and counting (Gongal et al., 2015).

Recently, modeling techniques incorporating vision-based machine learning, which have demonstrated success in numerous fields, have seen considerable use in fruit detection and counting via proximal imaging (Bargoti and Underwood, 2017; Gené-Mola et al., 2019; Santos et al., 2020). Much of the previous work has been performed in orchard crops, such as apples (Bargoti and Underwood, 2017; Häni and Roy, 2019; Wulfsohn et al., 2012), oranges (Maldonado and Barbosa, 2016), and mangos (Payne et al., 2014). Studies in other specialty crops, including tomatoes (Mu et al., 2020; Rahnemoonfar and Sheppard, 2017) and grapes (Liu et al., 2017; Liu and Whitty, 2015; Santos et al., 2020) have also been conducted. In vineyards, much of the existing literature in yield estimation from proximal imagery have consisted of methods leveraging feature engineering and computer vision to count individual grape berries (Millan et al., 2018; Nuske et al., 2014; Rose et al., 2016), although pixel count has also been related to grapevine yield (Diago et al., 2012).

In very recent years, studies on proximal imaging for fruit detection have moved away from hand-crafted feature and algorithm development towards deep learning methods (Gené-Mola et al., 2019; Milella et al., 2019; Santos et al., 2020). Deep learning methods used for fruit detection reduce the bias involved in model development via their flexibility in learning both the optimal features and optimal relationships between features to converge to the desired result. However, while these methods are robust to a small amount of occlusion, they cannot account for completely occluded fruits (Gongal et al., 2015; Mu et al., 2020). In an effort to overcome the issue of occlusion, previous research in grape yield estimation has demonstrated some benefit of incorporating variable grape visibility into yield estimation models for improving performance using statistical modeling (Millan et al., 2018; Nuske et al., 2014). However, these approaches have involved tuning the model to the specific dataset, either by incorporating previous years' data or specifying expected properties of the image data, such as mean berry area.

While previous yield estimation methods have focused primarily on semantic segmentation, object detection, and object masking used for fruit detection, deep learning methods have also been used to map image data directly to more abstract outputs, such as image captions, monocular depth estimates (Hu et al., 2019), scalar values of food calories (Ege and Yanai, 2017), subject age (Othmani et al., 2020), and more (Zakir Hossain et al., 2019). Once training is completed, these deep learning models map relationships from semantically complex combinations of image features to accurate predictions in unrelated modalities. This "end-to-end learning" approach can be applied to yield estimation by directly relating images and desired yield information, as opposed to the use of vegetative indices or models for detection or masking. This hypothesis was tested with promising results in the field of UAV-based yield estimation, in which a convolutional neural network (CNN) trained to forecast rice grain yield directly from high-resolution RGB images outperformed an NDVI-based approach (Yang et al., 2019). End-to-end learning has also been demonstrated in the context of grape yield estimation from proximal imagery, where a regression CNN model was used to directly predict yield in mass from images (Silver and Monga, 2019). However, images from that previous study were collected individually with a smartphone, which did not represent a scalable method. Additionally, vines were prepared specifically for imaging via placement of a calibration marker in each frame. Finally, extensive manual processing and cropping of the images was performed. Nevertheless, the study represents a proof-of-concept for utilization of DL in viticultural yield estimation. Like the improvements garnered by deep learning-based methods in learning features, as opposed to relying on prior assumptions, end-to-end methods used to directly relate images and yield may also be able to learn features from images other than solely those which contribute to fruit localization. For example, information such as canopy structure, fruit position, or other features may be relevant for the purpose of calibrating the yield estimate to account for occluded fruit. In addition to the above advantages garnered by end-to-end modeling, end-to-end models do not require labeling of objects in image data. Hand-labeling is a universal bottleneck in object detection studies and eliminating the labeling step can allow for the use of more data.

In this study, three different deep learning methods were compared in their ability to estimate yield in vineyards at varying levels of spatial resolution. Yield values in metric tons per hectare (t/ha) acquired during harvest were predicted using a fruit detection model as well as two end-to-end networks trained to predict yield directly from images: a CNN and a transformer network. The transformer architecture has recently been shown to perform well on a wide range of tasks, including natural language processing, as well as many vision-based tasks such as classification, segmentation, and object detection (Carion et al., 2020; Dosovitskiy et al., 2020; Vaswani et al., 2017; Xie et al., 2021). Considering the design of the transformer architecture in accepting sequences of data, such as words or image patches, this flexible architecture was applied to this dataset to allow for a set of neighboring images recorded near a ground truth yield measurement to be accounted for in the estimation.

In addition to demonstrating the application of novel modeling techniques, this study also incorporated a uniquely large dataset. Previous datasets used for model development have consisted of as few as 10 vines (Diago et al., 2012) to up to 1,212 vines (Nuske et al., 2014). Yield data collected by hand-weighing grapes is typically used as a ground truth measure. However, yield monitors which generate high-resolution yield data as crops are harvested are commonly used by growers to map yield on a large scale. These monitors have been used previously in conjunction with remote sensing studies for yield estimation in crops such as sorghum and cotton (Yang et al., 2004; Yang and Everitt, 2002) as well as grapes (Sun et al., 2017). However, proximal imaging studies on larger datasets of grape yield are still lacking, primarily due to the expensive nature of data collection. In this study, models were trained and evaluated on a dataset containing 23,581 yield values measured at harvest using a yield monitor. To the best of the authors' knowledge, this study is novel, as it represents the first application of end-to-end modeling using minimally processed images collected from a moving platform for grape yield estimation. Additionally, this study represents the first application of vision transformers in direct yield estimation from image data. Furthermore, the scale of the yield data considered in the present work was an order of magnitude greater than in previous studies. Finally, the data in this work were derived from a commercial, mechanically managed vineyard, as opposed to previous studies conducted in vineyards with manual management. Mechanized vineyards generally consist of increased occlusion, whereas manual management allows for better targeting of specific shoots and leaves as compared with machine implements.

5.3. Materials and Methods

5.3.1. Field Layout

Data were collected in a commercial grape vineyard (*Vitis vinifera* L. cv. Cabernet Sauvignon) within the California Central Valley region in September 2020. Image data were collected approximately three weeks before harvest, while ground truth data was collected during harvest. The vineyard was divided into four blocks representing differences in management and/or trellis type (Fig 5.1). For the purposes of this study, block was not considered during the training of any model. However, data are presented as separated by block to demonstrate the effect of management on model performance. Vines in most blocks were trained on a quadrilateral trellis with sprawling canopies, although one block contained grapevines with a single bilateral trellis. Grapevine rows were arranged in an East-West configuration for all blocks.

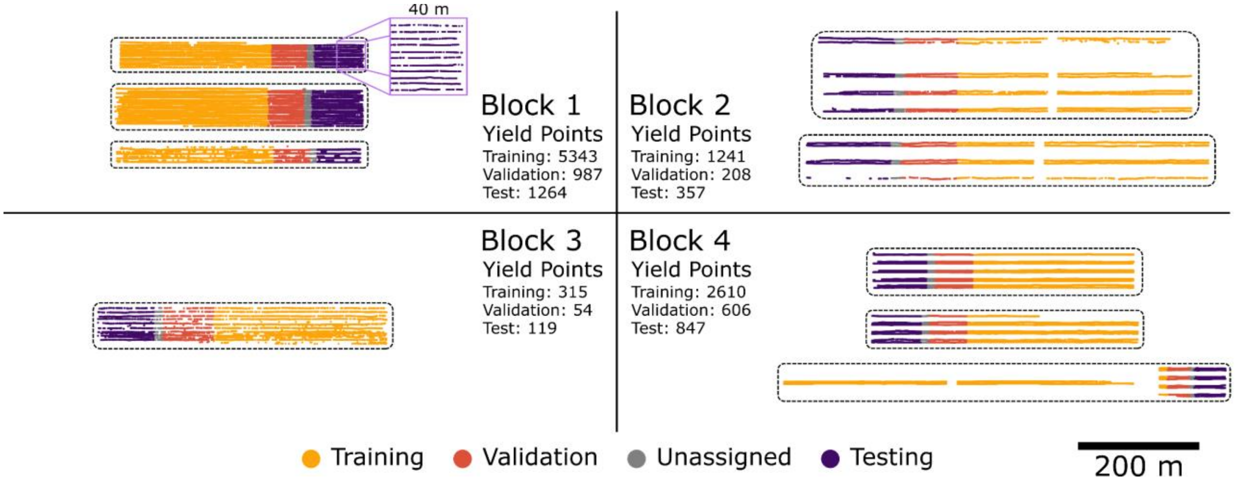


Figure 5.1. Yield points used in this study, colored according to their split in training, validation, and testing sets. Points located between validation and testing sets were labeled as unassigned and not used in the training process to minimize similarity between training and validation set images to the test set. The scale bar is applicable within each set of points bounded by a dashed line, but not between bounded regions, which have been positioned in the figure according to increasing block number for clarity.

5.3.2. Data Acquisition

5.3.2.1. Image Data

Image data acquisition was performed using a low-cost sensing kit with two RealSense D435i cameras (Intel, Santa Clara, CA) vertically oriented opposite each other, perpendicular to the direction of travel. Location data was collected using a Piksi Multi RTK GPS module and antenna (Swift Navigation, San Francisco, CA). In addition to the sensing modules, two 120W LED arrays on each side of the kit were included to illuminate the environment (Nilight, Englewood, NJ). All components were managed by a Jetson Xavier NX single board computer (NVIDIA, Santa Clara, CA) using the ROS platform. This sensing kit was attached to the back of an agricultural utility vehicle and powered by a deep cycle battery during data acquisition. Imaging was performed at night to control illumination consistency. RGB and depth image data were collected simultaneously from both cameras at 15 Hz frequency, although only the RGB imagery was used for this study. Images were compressed using JPEG compression to reduce storage requirements. Satellite-based augmentation system corrected GPS data was collected at 10 Hz, representing approximately 0.38 m spatial resolution along a row. Image georeferencing was performed based on timestamp matching.

5.3.2.2. Ground Truth Data

Grapes were mechanically harvested using a commercial harvester fit with a load cell along with a GPS unit for yield monitoring (Advanced Technology Viticulture, Adelaide, Australia). Force data from the load cell were recorded at 1 Hz and the harvester was driven at an average speed of ~1 m/s. Data were collected continuously and the yield monitor was calibrated with a scalar time offset such that the location recorded for each yield value best represented the location where the grapes were grown.

5.3.3. Data Handling

5.3.3.1. Ground Truth Data

Yield monitor data was first processed manually to remove outliers, defined as yield points from gaps in rows where recorded yield was artificially lowered, spurious GPS points from outside of the rows, and rows in which the frequency of yield points was erroneously low. Values were then filtered further to remove points more than 1.5 interquartile range values away from the first and third quartiles (Tukey, 1977). Next, data from the yield monitor was calibrated based on the total yield from each vineyard as weighed by the receiving winery. Yield density was calculated by dividing the mass of grapes per block (tons) by the area of each block (hectares). All yield points within each block were then scaled proportionally such that the mean value was equivalent to the mean yield density measured in each block using a commercial scale as part of routine operations.

5.3.3.2. Image Data

The total ground-imagery dataset before filtering and data association consisted of 274,944 images. Image data were filtered to remove poor quality images, which consisted of images where a shoot extending into the inter-row space obscured the camera, images recorded during turns between rows, and excessively blurred images where individual grape clusters could not be seen clearly. This was done by randomly selecting 3,000 images from the dataset and manually labeling them with binary labels (“keep” and “toss”). This subset of labeled data was split into training, validation, and test sets with 1800, 600, and 600 images, respectively, and a

MobileNetV2 model was trained to filter images (Sandler et al., 2018). The trained model achieved an accuracy of 91.2% on the test set and was used to filter the rest of the dataset.

5.3.3.3. Association of Image and Yield Monitor Data

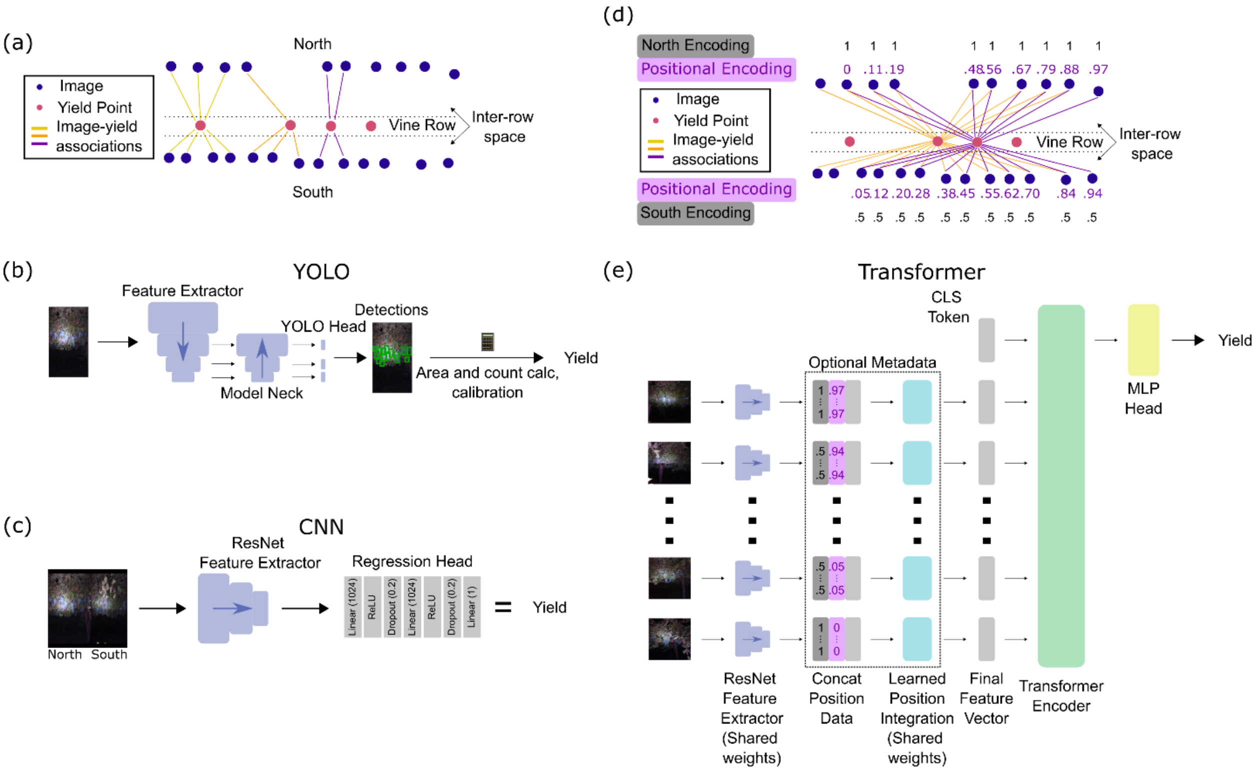


Figure 5.2. Data handling and model architectures. a) Images in object detection and CNN models were associated with their closest yield point (three yield points shown with image-yield associations, where the images associated with each point are shown in different colors). b) The YOLO model was applied to individual frames and detections were processed to calculate yield values at each point. YOLOv5 was used without modification (Jocher et al., 2021b). *c) The CNN architecture was applied to a pair of images, one from each side of the vine. The model contained an unmodified ResNet feature extractor with a custom set of linear layers added in series. The final linear output of the model represented the predicted yield. No softmax or other post-processing was performed. d) The transformer model was trained to associate a 10-meter window of image points with each yield point, such that images may be associated with multiple yield points (two yield points shown with image-yield associations in different colors). Positional information consisting of positional and orientation (north and south) encodings was extracted based on relative location of image and yield points. e) The transformer architecture was designed to accept any number of images up to 128 (first and last two shown). Each image was passed through the same ResNet feature extractor (weights were shared during training). When used, positional encoding was represented as vectors with the scalar encoded value repeated such that a vector the same size as the feature vector from the ResNet extractor was produced. Positional encoding was integrated using a learned weighted average for each image input (weights were shared during training). The transformer encoder was composed with a depth of 2 layers and 8 attention heads. The class token was used to generate a prediction, as it represented the entire input sequence.*

Yield data was collected during harvest, and each yield point was marked directly over the grapevine row, with a different spatial resolution as compared with image data. Therefore, association of yield and image data was necessary (Fig. 5.2). For association of each image with a corresponding yield measurement, a distance matching algorithm was designed to match each image with its closest yield point, considering the orientation of the camera (North or South) for each image. For training the transformer model (Section 5.3.4.3), all measured yield points with at least one North-facing and at least one South-facing image within 5 m to the east or west of the measured yield point were kept. Any yield point without images from both the North and South side of the vine were discarded. This set of measured yield points was further pruned for training the CNN model (Section 5.3.4.2), where yield points were only retained if they were associated with at least one image from each side of the vine that was not closer to an adjacent yield point (Fig. 5.2a). In total, this left 23,581 yield points in the dataset used to train the transformer model, and 14,302 yield points in the dataset used to train the CNN model. After removing poor quality images, images far away from yield points, and images associated with yield points for which both sides of the vine were not accounted for, these yield points were associated with 164,699 and 80,009 images in the transformer and CNN model sets, respectively.

5.3.3.4. Dataset splitting for model development

Yield values were split into training, validation, and testing sets according to their location, such that no training or validation data was adjacent to a point in the test set. This was performed to reduce the influence of spatial autocorrelation on model performance (Fig. 5.1).

Within the dataset used to train the transformer model, this resulted in 15,024 yield points in the training set, 3,354 in the validation set, and 4,529 in the test set. For the CNN and object detection models, this resulted in 9,509 yield points in the training set, 1,855 in the validation set, and 2,587 in the testing set.

5.3.4. Model Architectures and Training Procedures

5.3.4.1. Object Detection

The object detection model required labeled bounding boxes around grape clusters visible within a subset of images. Specifically, 150 images were selected at random from the training and validation sets of the dataset. Grape bunches in these images were labeled with bounding box labels. Care was taken to only label bunches on the near side of the vine, in the event that bunches from the far side of the canopy were visible in the image. For vines trained with quadrilateral trellises, grapes on the near side were defined as grapes on the closer set of cordons. For vines trained with bilateral trellises, the cordons were used as a dividing plane, where grape bunches on the near side of the cordons were selected. These 150 images were divided into training, validation, and testing sets of 98, 24, and 28 images respectively, considering the distribution of images from each management block.

The object detection model was trained with YOLOv5 (Jocher et al., 2021b). Images were resized to 640 x 640 pixels and augmentation was performed using the mosaic method (Jocher et al., 2021a), along with variation of hue, saturation, and luminance values representing the default augmentations. The model was trained for 300 epochs. The best model was determined using a weighted average of a) 0.9 times mean average precision (mAP) averaged over intersection over union (IOU) values of 0.5:0.95 in increments of 0.05 and b) 0.1 times mAP at an IOU of 0.5 only. This evaluation metric was computed on the validation set after each epoch.

Bounding box count (i.e. bunch count) and bounding box area per image were used in yield estimation. For each location with yield data, mean bunch count or mean summed bounding box area was determined for all images on each side of the vine. The two sides were then summed to obtain the final estimate of bunch count or area (px^2). Bunch count and area were then correlated with the yield values in all training set images from the full dataset. A simple linear regression with the intercept fixed at the origin was used to relate bunch count and area with yield. Evaluation on the test set was conducted by first obtaining mean count or summed area over both sides of the vine, then converting the value to yield with the corresponding linear fit.

5.3.4.2. Convolutional Neural Network Model Architecture

For the CNN model, the ResNet18 architecture (He et al., 2016) was adapted to a regression approach (Fig. 5.3a). The 1000-length final linear layer of the architecture was replaced with a regression head composed of two 1024-length linear layers, each followed by a ReLU and 20% dropout step (Nair and Hinton, 2010; Srivastava et al., 2014). A final linear layer of size 1 was used to represent the yield corresponding with the input. Weights for the model were randomly initialized. Input data consisted of one frame from each side of the vine concatenated horizontally (Fig 5.2c). Since multiple frames on either side of the row were associated with a given yield point, the North side and South side image used in the input were randomly selected from the available images associated with each point during training. The validation set consisted of seeded random selection of frames, such that the North and South frames were consistent between epochs.

The model was trained using an adaptive loss function (Barron, 2019) for 1.97M steps of batch size 12, which were divided into 25 epochs. North and South frames were augmented separately using random horizontal flips, and random median blurring. The North image was always on the left side of the input. For inference on the test set, all combinations of North and South image pairs were input to the model for each yield point, and the average predicted yield was taken.

Model weights were saved and a validation score was computed after each epoch. The model with the lowest validation loss was selected for performance evaluation.

5.3.4.3. Transformer Model Architecture

While concatenation of two images was used to increase the context of the data with respect to the location at which yield data were measured by providing two views of the vine in the CNN approach, concatenating additional frames would quickly exceed available GPU memory or require substantial image downscaling leading to information loss. The vision transformer architecture employed in this study accounts for this limitation by accepting a sequence of images as its input. Additionally, due to the attention mechanisms in the model architecture, the influence of each input image to the final predicted value is allowed to vary (Dosovitskiy et al., 2020; Vaswani et al., 2017). This was of particular importance in this work, due to the use of a mechanical yield monitor in lieu of hand counting or weighing fruit. While the yield monitor allowed for generation of a uniquely large dataset, both the magnitude as well as the positional accuracy of the yield measurements were compromised by the continuous nature of data collection. Specifically, harvested grapes were measured with a load cell after a small delay, during which the grapes were conveyed from the vine to the instrument. While this time was accounted for with a scalar offset in this study, other factors such as the mass flow rate of fruit in and out of the harvester machinery have been shown to affect this value slightly (Searcy et al., 1989). As such, the yield values reported in this study were more appropriately derived from a distribution of vines surrounding the marked position. Therefore, a model with the capacity to accept data from areas surrounding the location of the measured yield point was particularly well-suited to the present task.

To take advantage of the flexibility of the transformer architecture, a window of 5 m to the east and west of each yield point was considered, and all images within that window were used as inputs to the model for the purpose of predicting yield in that location (Fig. 5.2d). This was done to account for any mixing which may have occurred in the harvester as yield was being recorded, potentially allowing adjacent vines to influence the recorded yield in a given location.

The transformer model was based on the Vision Transformer (ViT) architecture (Dosovitskiy et al., 2020) with modifications (Fig. 5.2e). Instead of linearized image patches, token vectors input to the encoder were represented by ResNet34 features. Each image passed through the ResNet model, and the final set of activation maps was used for feature extraction. Maps were averaged along the filter dimension, then linearized to generate a 256-length feature vector for each image.

In addition to the token representation, the positional encoding was also modified from its original format. Typically, positional encoding consists of sinusoidal or learned embeddings (Vaswani et al., 2017), which have been shown to be effective for inputs derived from uniformly cut patches of an image or language inputs. However, for both patch and word-based input, the spacing between inputs is consistent and only the relative location between inputs is important. In this application, the relevant positional information consisted of the distance between the measured yield point and the image location, as well as the orientation of the camera with respect to the vine. As a result, the prior information regarding the distance between each frame and the yield point, as well as the side of the vine represented by each image was used as the positional embedding. Setting the location of yield point at 0.5, locations 5 m to the east and west were scaled to 0 and 1, respectively. This scalar was assigned to each input image and represented its position relative to the center frame at 0.5. Likewise, each input image was assigned either 0.5 to represent an image from the South side of the vine, or 1 to represent the north side. Finally, as opposed to the typical summation of a positional vector with the token vector, a 1D convolutional layer was added to allow for a linear combination of each value in the 256-length feature vector with each of the scalar positional values to be learned and used to improve performance. The inclusion of this positional metadata was performed to allow the model to selectively attend to images based on both their content as well as their position. To determine the influence that this added positional information had in an agricultural system, the model was trained both with and without information regarding position and orientation of each image frame (this optional step is represented in a dashed box in Fig. 5.2).

The remainder of the model architecture was minimally changed from the ViT architecture. The encoder was designed with a depth of 2 encoding blocks and 8 attention heads. The multilayer perceptron decoder used for classification was modified to output a single value to serve as the scalar prediction of yield. The model was trained with a mean squared error objective function for 50,000 steps and a batch size of 6, with gradient accumulation across every two batches for an effective batch size of 12. Training was divided into 20 epochs. Model weights were saved and validation performance was measured after each epoch. The weights with the lowest validation error were kept for performance evaluation.

5.3.5. Performance Evaluation

Object detection performance within the labeled dataset was measured using area under the precision-recall curve (AP) at an intersection over union of 0.5, as well as R^2 and root mean squared error (RMSE) between labeled and predicted bunch counts and bounding box area.

Performance of each model on the test set was evaluated using RMSE, mean absolute percent error (MAPE), and R^2 metrics between predicted and measured data. In order to compare all models, only the yield points represented in the test set of the CNN model were used for evaluation (Section 5.3.3.3). Therefore, all models were evaluated on a test set of 2,737 yield points. Additionally, due to the tendency for some of the trained models to predict values closer to the mean, as opposed to higher and lower yield values seen less frequently, the range expressed by each approach was also calculated using Eqn. 5.1:

$$\frac{\max(predicted\ yield) - \min\ (predicted\ yield)}{\max(measured\ yield) - \min\ (measured\ yield)}\ x\ 100 \qquad (5.1)$$

Finally, to compare performance with remote sensing-based studies, post-hoc analysis was performed after model inference by spatially aggregating yield points and associated ground truth and predicted data into 10- or 20-meter square bins within each block. This was done by dividing the test dataset into separate zones that were densely populated with yield points, then binning points in the zones into grids of either 10- or 20-meter length, beginning at the lower left point of each zone. Within the test set, 179 bins were used at 10 m spacing, and 47 bins were used at 20 m spacing.

5.3.6. Saliency Mapping and Model Visualization

To gain insight into the features of the image used by the model to predict yield, the technique of Gradient-Based Class-Activation Mapping (Grad-CAM) was leveraged for the CNN model (Selvaraju et al., 2016). Typically used for classification tasks, this technique allows for visualization of the regions of the input image which contribute most strongly to increasing the value of an output class. As the model only outputs one value, the predicted yield, the visualization can be interpreted as the regions of the input image which contribute to raising the predicted yield value. While these regions are hypothesized to be localized in regions of input images with visible grape bunches, other features of the input, such as canopy density or shoot position, may also be relevant.

The Grad-CAM approach was used on all images in the train set as well as the test set (all combinations of images from each side of the vine) to produce heatmaps scaled from 0 to 1. These maps were then averaged within each yield point, then averaged again over the entire dataset, giving equal weight to all yield points regardless of the number of associated images. For comparison with the Grad-CAM visualization, a heatmap was generated in a similar fashion for the train and test sets of the object detection model, where detected clusters in each image were assigned a value of 1 on a blank canvas of a size equal to the image size. These images, like the Grad-CAM heatmaps, were averaged within each yield point separately for the north and south sides of the vine, then averaged over all yield points to generate a similar heatmap to the CNN Grad-CAM plot.

Finally, in addition to visualization of model outputs, the scale of the dataset in this work was sufficient to produce yield maps. Maps were generated by aggregating yield points spatially at 10 m resolution, then plotting the aggregated regions on a map colored by yield value.

5.4. Results and Discussion

5.4.1. Object Detection Approach

5.4.1.1. Internal Validation

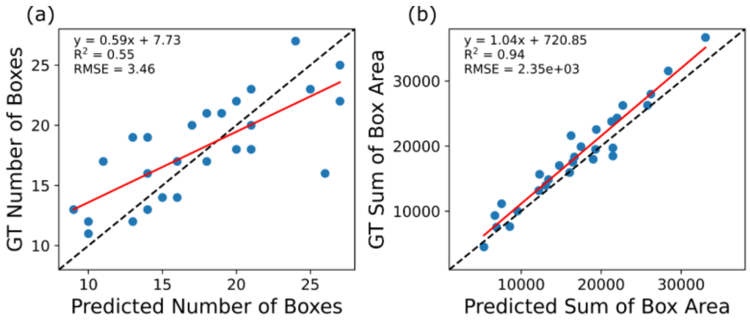


Figure 5.3. Internal validation of YOLOv5 model on A) box count and B) summed box area hand-labeled test set of 28 images. The red line represents the linear fit to the data and the dashed black line represents 1:1 accordance.

Within the 28 labeled test images, the model achieved an AP score of 0.56, an R^2 of 0.55, and a RMSE of 3.5 bunches in prediction of the number of grape bunches in each image (Fig 5.3). Prediction of summed bounding box area demonstrated an R^2 of 0.94.

While many previous studies have focused on berry counting, Santos et al. (Santos et al., 2020) previously demonstrated an application of CNNs in grape detection in vineyards, including the YOLOv3 object detection and Mask-RCNN segmentation models. The AP score achieved by the researchers for YOLOv3 at 0.5 IOU was 0.39, which is considerably lower than the YOLOv5 score in this study, although the YOLOv2 model in the previous work achieved a score of 0.48, which is similar to this study. In this previous study, Mask-RCNN achieved an AP of 0.72, representing a considerable improvement. However, Mask R-CNN is a much more computationally expensive network than YOLOv5.

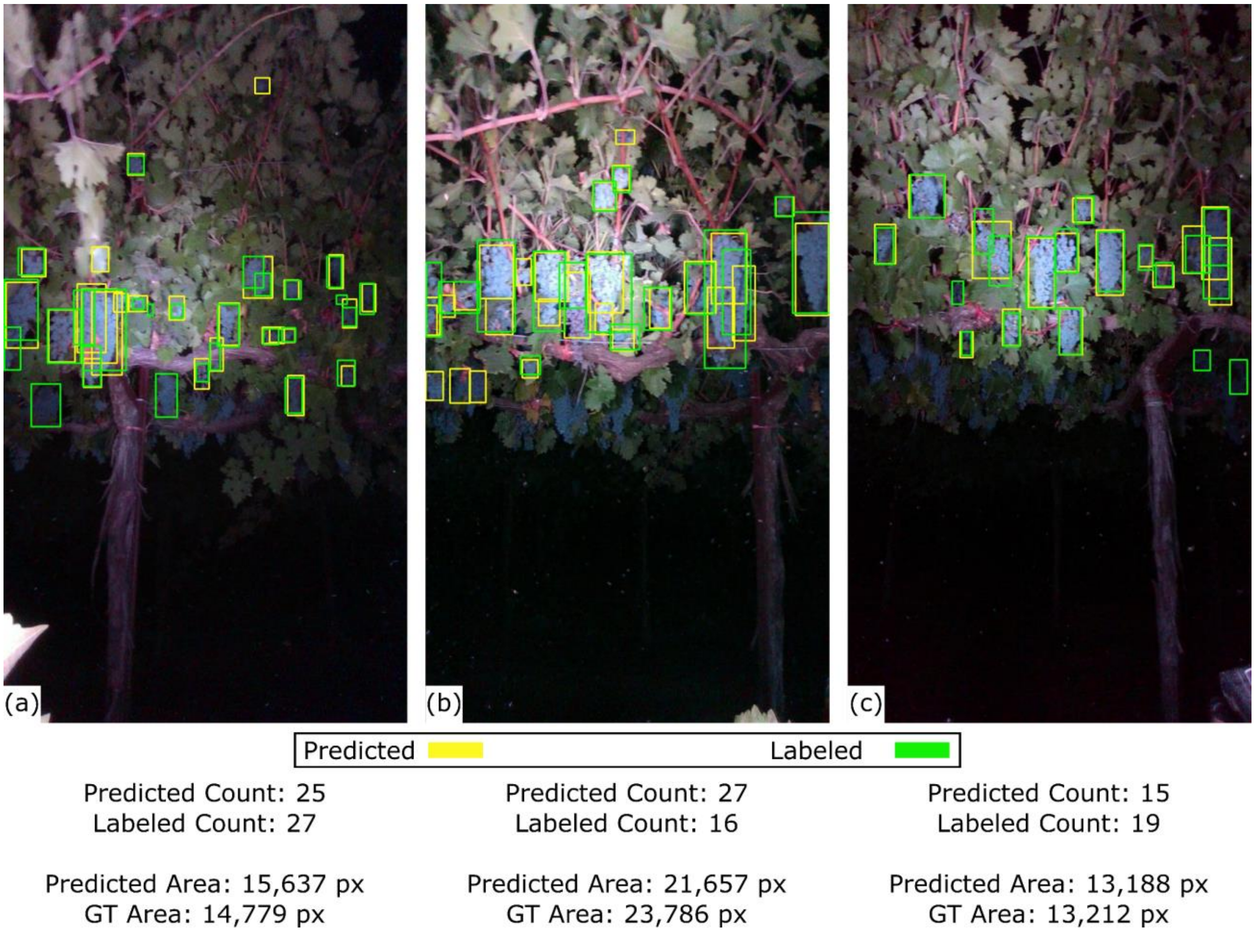


Figure 5.4: Example images with predicted (yellow) and labeled (green) grape bunches. Only bunches on the near side of the vine were labeled. Quantitative measures of grape yield in bunch count and area are displayed below each frame. a) In this frame, missed bunches in the lower-left as well as additional bunches near the top of the frame are demonstrated relative to ground truth label. b) In this frame, the predicted count is much higher than the labeled count, but the predicted area is approximately equal to the labeled area. c) Likewise, in this frame, the predicted count is less than the labeled count, but the predicted area is also very close to the labeled area.

While the AP metric is a valuable indicator of model performance, it may not represent performance relevant to yield estimation. Notably, inconsistencies in bunch counts amounted to an RMSE of 3.46 bunches, primarily due to overcounting in the model. However, the degree of counting error was inconsistent, leading to an R^2 of 0.55. This inconsistency may have been due to clustering of bunches, resulting in multiple bunches classified as one, which is a common issue (Di Gennaro et al., 2019; Liu et al., 2017). Moreover, additional grape bunches missed during labeling were detected by the model in some instances (Fig 5.4). These bunches may have constituted those on the far side of the vine (Section 5.3.4.1), such as in Fig. 5.4b. However, although the raw value of bunch counts was not always accurately predicted, the correlation between measured and predicted summed box area is strong, with an R^2 of 0.94 (Fig 5.3b). With an area-based approach, two labeled bunches counted as one as well as one labeled bunch split into multiple individual bunches may not influence the predicted area, even when the count is affected. These results suggest that in the case of box area, the object detection model is consistent with human-labeled fruit annotations.

5.4.1.2. Yield Estimation using Object Detection

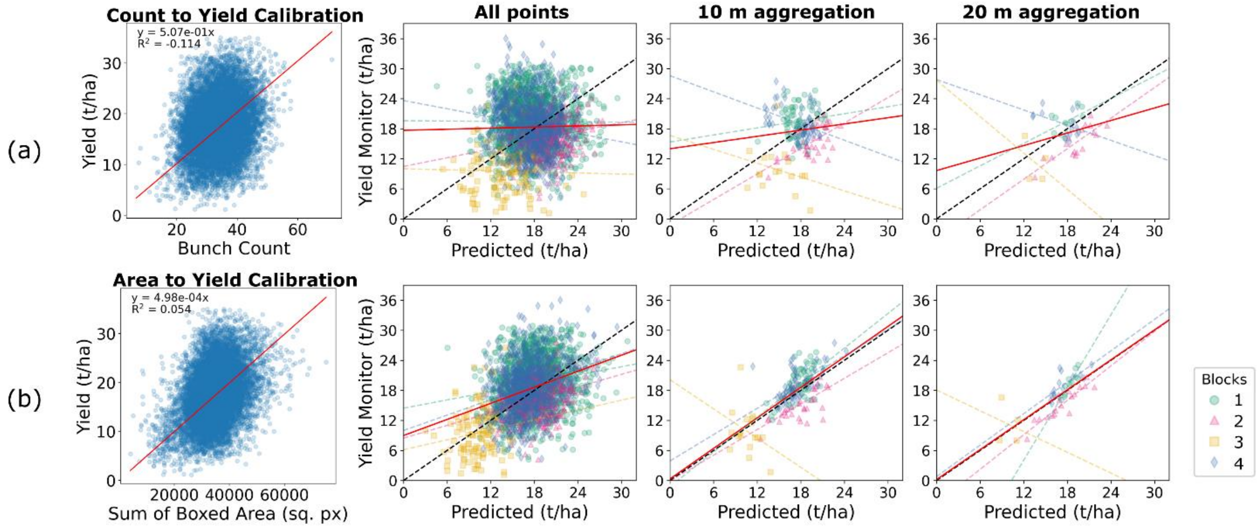


Figure 5.5. Yield estimation on data from the test set using the object detection model with both a) box count and b) summed box area approach. The first pane represents the calibration performed using the training set data to translate either bunch count or box area to ground truth yield value in tons per hectare (t/ha). The red line in each pane represents the fit linear model to all data points in the pane. Each dashed line represents the model fit to each block of corresponding color. Black dotted lines represent the 1:1 line. The level of spatial aggregation is shown above each column after the calibration results (first column).

To predict yield with the object detection model, correlations were made between bunch count and ground truth yield as well as boxed area and ground truth yield (Fig. 5.5). Using the training set, the linear relationship between count and t/ha demonstrated a poor fit, with an R^2 of -0.11. Consequently, use of the bunch count model with this relationship showed poor accordance with yield in the test set with a RMSE of 6.27 t/ha and MAPE of 34% (Table 1). This result is similar to previous work, as studies incorporating bunch detection steps have typically used the bunch counts along with berry counts due to the variance in size between bunches (De La Fuente et al., 2015; Nuske et al., 2014b) or used pixel area to express the difference between bunch sizes (Di Gennaro et al., 2019).

Table 5.1. Performance summary for each architecture on test set data only. Between the three levels of spatial aggregation, "None" is represented by 2587 yield points, 10 m by 179 yield points, and 20 m by 47 yield points.

Architecture	**Details**	**Spatial Aggregation (m)**	**RMSE (t/ha)**	**MAPE (%)**	**R^2**	**Range Expressed (%)**	**Fit Line Slope**	**Fit Line Intercept (t/ha)**
Object Detection	Bunch Count	None	6.27	34.0	0	71.3	0.04	17.7
		10	4.75	29.4	0.013	49.7	0.21	14.0
		20	4.00	22.1	0.081	74.5	0.42	9.64
	Box Area	None	5.39	27.5	0.117	79.7	0.53	8.98
		10	3.32	18.5	0.446	62.2	1.01	0.32
		20	2.46	12.2	0.577	86.3	1.0	0.17
CNN	--	None	5.06	26.9	0.136	44.7	0.87	2.81
		10	3.16	18.7	0.510	49.0	1.3	-5.15
		20	2.78	15.1	0.501	63.2	1.2	-4.13
Transformer	No Metadata	None	5.26	28.0	0.096	56.5	0.61	7.37
		10	3.47	19.6	0.376	57.7	0.97	0.54
		20	3.07	16.5	0.371	79.9	0.97	-0.20
	Position, Orientation Information	None	5.12	27.1	0.149	60.0	0.64	6.72
		10	3.24	18.0	0.460	63.8	0.91	1.73
		20	2.89	15.2	0.430	80.5	0.85	2.42

Like in the bunch count approach, yield estimates using bounding box area require a relationship to be mapped between area and yield on the training set. This linear fit performed better than the bunch count results, but was still a poor fit overall, with an R^2 of 0.054 (Fig. 5.5b). This value is still lower than previously reported correlations between pixel count and grape yield, with Diago et al. (2012) reporting an R^2 of 0.76. However, the field layout used in this previous study was comparatively simple, with a white background placed behind each vine to avoid influence from adjacent rows, increased image resolution, and a vertical shoot positioned (VSP) trellis system in which cluster occlusion was reduced. In comparison, the quadrilateral cane trellis system featured primarily in this work exhibited increased occlusion, and the images were collected without compensation for adjacent rows (see Section 5.4.3 below).

Interestingly, although the relationship between summed area and yield is poor, the internal validation results demonstrated good accordance with labeled data (Fig. 5.3b), demonstrating that the relationship between observed bunches and yield value is inconsistent. Detection of grape bunches on the far side of the vine may have influenced the yield estimation, although this was not the case in the labeled test set (Fig. 5.3b). Therefore, the results indicated that even if a model were to accurately label all grape bunches as well as a human, the relationship to yield may still be poor. This suggests the amount of fruit present on the vine but invisible to the camera lens was not consistent throughout rows, as is sometimes assumed for the sake of modeling occlusion (Bargoti and Underwood, 2017).

Still, even with poor performance in relating area to yield, for predictions on the test set, performance using the box area approach was improved over the bunch count approach at all levels of spatial aggregation. Performance at the 10 m level aggregation demonstrated an R^2 of 0.45 and RMSE of 3.32 t/ha. This increase in performance over the bunch count model was most likely due to accounting for variability in bunch size. Additionally, while the bunch count model was only able to express 49.7% of the range of values at 10 m aggregation, the box area model increased this to 62.2%. Moreover, looking at the slope and intercept of the linear fit to the test data (Table 1), the relationship is close to a 1:1 accordance, with a slope of 1.01 and intercept of 0.32 t/ha.

In addition to the high levels of occlusion in vines imaged in this study relative to previous works, one potential reason for decreased performance relative to existing object detection approaches is the small number of labeled images. Only 150 images were labeled, with 98 images used to train the model, representing 2630 total labeled bunches, and 1696 labeled bunches in the training set. For comparison, the Wine Grape Instance Segmentation Dataset (WGISD) published by Santos et al. (2020) contains 300 images with 4432 clusters. However, the dataset in the present work was distinct as the vineyards were mechanically managed, and therefore the occlusion level was much higher. Additionally, the WGISD was created as an instance segmentation dataset, as opposed to the object detection dataset used in this work. Even so, labeling images for the present study represented a considerable amount of labor as well as a bottleneck in model development. This has been noted in previous works (Rahnemoonfar and Sheppard, 2017; Santos et al., 2020), in which the laborious nature of image labeling was specifically noted. This labeling effort becomes increasingly burdensome as the variability of images conditions increases, pointing to a need for more data efficient methods for labeling, such as in Fei et al. (2021). In this work, the requirement for labeling was sidestepped in the regression approach.

5.4.2. Regression Approaches

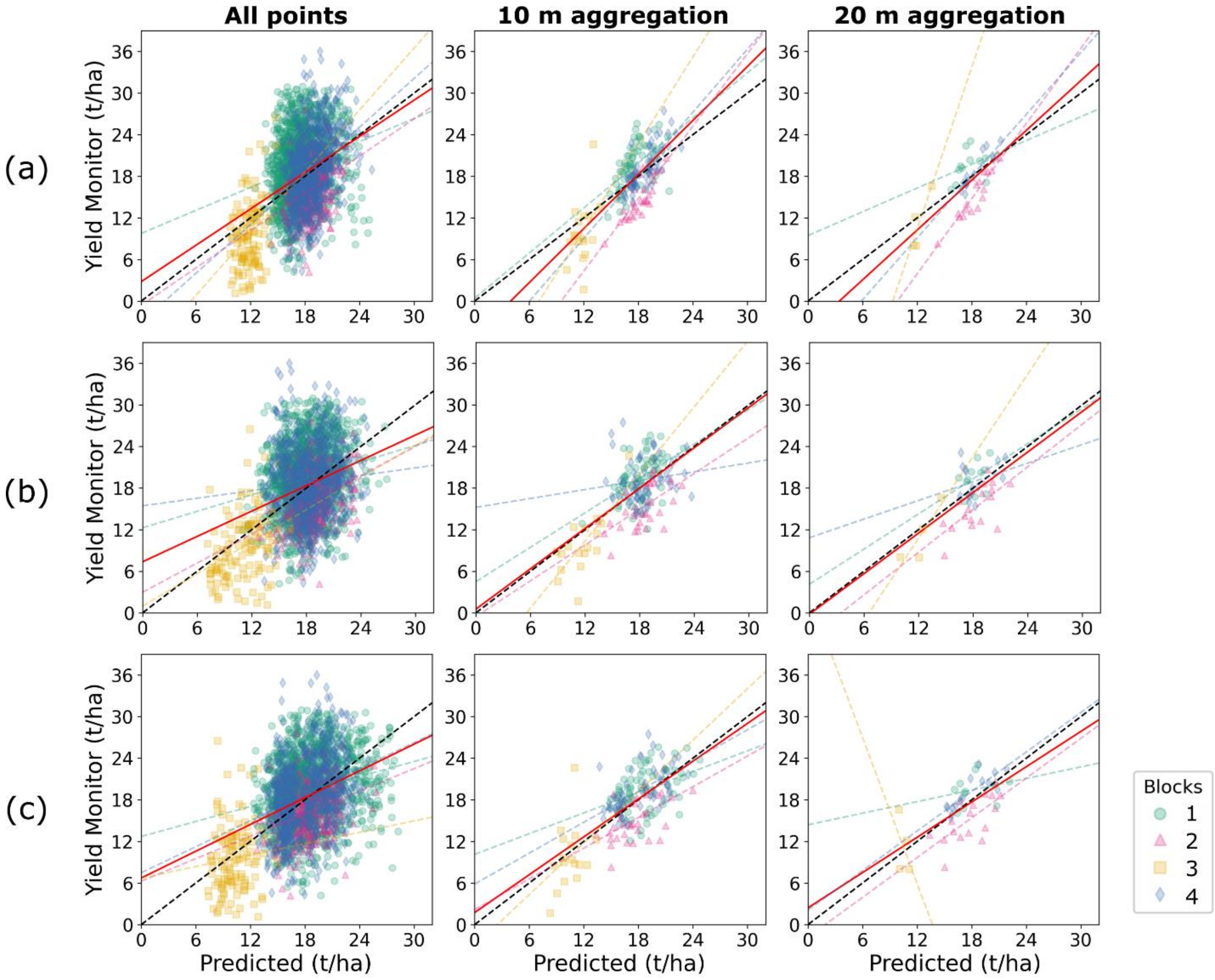


Figure 5.6. Yield estimates on test data using the regression models: a) CNN regression network; b) transformer model without positional; and c) transformer model with positional information encoding frame position relative to yield point (scaled from 0 - 1) and orientation (North/South). The red line in each pane represents the fit linear model to all data points in the pane (fit parameters shown in Table 1). Each dashed line represents the model fit to each block of corresponding color. Black dotted lines represent the 1:1 line. Each pane represents an increasing amount of spatial aggregation.

Unlike the object detection approach, each of the regression models were trained end-to-end on yield from images as input, which resulted in an output of yield in units of t/ha. Additionally, no manual labeling was required after alignment of yield points with image locations (Fig. 5.2). Relative to the object detection models, the CNN model further increased performance in yield estimation at 10 m spatial aggregation (Fig. 5.6a), with an RMSE of 3.16 t/ha, MAPE of 18.7%, and R^2 of 0.51 (Table 1). However, the model was only able to express 49% of the measured range at 10 m aggregation. Additionally, the line of best fit demonstrates bias, with a slope of 1.3 and intercept of -5.15 t/ha.

The transformer model without included metadata demonstrated similar performance to the CNN model (Fig. 5.6b), with an R^2 of 0.38, MAPE of 19.6%, and RMSE of 3.47 t/ha at 10 m aggregation (Table 1). While the R^2 and other error metrics were lower, the line of best fit represented close to 1:1 accordance, with a slope of 0.97 and an intercept of 0.54 t/ha. Additionally, range expressed by the transformer model demonstrated a slight improvement over the CNN approach, with 58% vs. 49% in the CNN. This flexibility to predict values at the extents of the measured range may be a result of the increased context of the input data combined with the attention mechanism of the model architecture. However, it is notable that the box area model achieved an increased level of range expressed compared to the CNN and transformer model without positional metadata, with the box area model showing 62.2% range expressed compared with 49% and 58% expressed by the CNN and transformer models, respectively.

However, when positional metadata was added to the transformer input, range expressed as well as MAPE was improved over all other architectures (Fig. 5.6c), with an R^2 of 0.46, MAPE of 18.0%, and RMSE of 3.24 t/ha at 10 m aggregation (Table 1). However, the line of fit was slightly further away from 1:1, with a slope of 0.91 and intercept of 1.73 t/ha. However, the range expressed by the model increased over the other end-to-end approaches, with 64% expressed at 10 m aggregation. Since the only modification relative to the other transformer model was the addition of positional metadata, this data likely allowed the model to better attend to the input data and selectively map the inputs to higher and lower yield values. The addition of positional metadata to deep learning models in the agricultural domain has been previously demonstrated by Bargoti and Underwood (2015, 2017) in the study of image segmentation in apple orchards. Along with input image patches, these previous models were designed to accept metadata including pixel positions, row numbers, and solar position. This additional information was shown to improve performance of a multi-scale multi-layered perceptron (MLP) segmentation model, but negligibly impacted performance of a CNN model. This may have been due to the interaction between the added data and the model architectures. In the present work, the metadata included with the transformer model was added before the transformer encoder layers, as opposed to the CNN model in the previous work, which added metadata in one of the last layers (Bargoti and Underwood, 2017).

Of all models, the best performance by MAPE (as well as range expressed) at 10 m aggregation was achieved by the transformer model with positional metadata, with an MAPE of 18% and 63.8% range expressed. However, at 20 m aggregation, the box area model achieved the best results, with a MAPE of 12%. Range expressed by the box area model was also the highest of all models at 20 m aggregation, with 86%. However, it should be noted that the box area model could only be trained after labeling 150 images of grape clusters, whereas the end-to-end models required no labeling. Moreover, the architecture of the end-to-end model is flexible towards the addition of metadata, which increases the potential for integration of proximal imagery models with other data sources, such as remote sensing data, for the improvement of predictive performance.

5.4.3. Performance Relative to Previous Works

The model performance at varying spatial aggregation values can be compared with previous remote sensing studies, such as Sun et al. (2017), in which vineyard yield in a location in the

California Central Valley, similar to the one in this work was predicted at 30 m using NDVI and LAI. In their work, they achieved up to 5.9 – 14.8% error, although the authors did not split the data into training and testing sets, so the error is correlation error instead of prediction error. Additionally, that value represented the best possible correlation among many possible combinations of cumulative vegetation index maps created across the season, which the authors noted cannot be known a priori.

At 20 m spatial aggregation (selected as the most similar aggregation level to 30 m satellite models), the regression models in this study performed well, with the transformer model with positional metadata demonstrating 2.89 t/ha RMSE and 15% error. However, performance evaluated without spatial aggregation was worse, with an RMSE of 5.12 t/ha and 27.1% error (Table 1). The values achieved without spatial aggregation are considerably lower than those of previous works, with Nuske et al. (2014b) demonstrating a relationship between detected berry count and yield with an R^2 of between 0.6 and 0.73 on an individual vine level. Other authors have obtained similar results with vine-level relationships with R^2 values > 0.7 (Diago et al., 2012; Millan et al., 2018). However, previous studies on yield estimation have focused on unsupervised computer vision techniques, such as use of keypoint detection (Nuske et al., 2014b) or distance-based metrics performed on color data (Millan et al., 2018). In these previous approaches, models were developed based on the appearance of images which were also used to generate performance metrics, as opposed to this study, which implemented a representative holdout test set. For example, one previous study on yield estimation from UAV imagery of grape canopies used images from one harvest year to estimate yield in the following year. However, according to the authors, images used for performance evaluation were selected as those with the best conditions, as opposed to a representative sample (Di Gennaro et al., 2019).

Notably, the previous end-to-end modeling study for grape yield estimation evaluated performance using 5-fold cross-validation between the 40 vines included in the study (Silver and Monga, 2019). However, the authors did not account for similarities in vines due to proximity in space. Therefore, a better comparison with correlative remote sensing as well as unsupervised proximal imaging approaches may be the performance achieved on data used to train the model (Table 2). In that context, models in the current work perform very well, with the transformer model with positional metadata achieving an R^2 of 0.91 at 20 m spatial aggregation. Notably, however, results without spatial aggregation are still low, with an R^2 of 0.54 in the same transformer model.

Table 5.2. Performance summary for each architecture on training set data only. Between the three levels of spatial aggregation, "None" is represented by 9509 yield points for the object detection and CNN models and 15024 yield points for the transformer models. 10 m is represented by 577 yield points for the object detection and CNN models, and by 617 for the transformer models. Finally, 20 m is represented by 171 yield points for the object detection and CNN models, and 182 points for the transformer models.

Architecture	Details	Spatial Aggregation (m)	RMSE (t/ha)	MAPE (%)	R^2	Range Expressed (%)	Fit Line Slope	Fit Line Intercept (t/ha)
Object Detection	Bunch Count	None	6.12	34.2	0.044	96.7	0.35	11.65
		10	4.05	20.3	0.242	58.3	0.79	4.17
		20	3.62	16.8	0.254	57.2	0.66	6.62
	Box Area	None	5.64	30.7	0.129	104	0.57	7.75
		10	3.36	15.5	0.477	74.6	0.95	1.41
		20	2.74	12.4	0.520	79.9	0.91	1.91
CNN	--	None	4.73	27.2	0.342	55.2	1.15	-2.59
		10	2.49	12.5	0.732	55.0	1.26	-4.63
		20	2.01	9.2	0.761	59.0	1.22	-3.99
Transformer	No Metadata	None	3.93	23.1	0.568	84.4	1.04	-0.71
		10	1.40	6.3	0.908	87.3	1.06	-0.86
		20	1.13	4.8	0.932	86.2	1.07	-0.90
	Position, Orientation Information	None	4.09	23.4	0.535	95.3	0.93	1.15
		10	1.49	6.4	0.893	87.8	0.97	0.67
		20	1.23	4.9	0.910	90.7	0.99	0.49

However, in addition to the lack of a holdout set, previous studies were conducted in very different environmental conditions as compared with this work. Almost all previous studies on grape yield estimation have been performed with VSP trellis configurations and manual defoliation. As an example of the effect that trellis configuration has on proximal imaging, one previous study of image-based methods for grape phenotyping noted that in point cloud classification of grape pixels, precision and recall dropped by 39% and 6%, respectively, between vines trained with a VSP trellis and semi-manual pruned hedge (SMPH) trellis (Rose et al., 2016). The SMPH trellis configuration seen in that work is more similar to the quadrilateral trellis featured primarily in the present study as opposed to the VSP trellis. Additionally, Di Gennaro et al. (2019) separated vine images into conditions based on vigor as well as occlusion and found the true positive rate of pixel classification was reduced by 45 and 73% in high and low vigor vines, respectively, between good conditions representing minimal occlusion and poor conditions representing occlusion and shading. On the contrary, this work was performed in a commercial vineyard within the California Central Valley, where due to the heat and dry conditions, fruit shading is extremely important, and VSP trellis configurations and defoliation would be detrimental to grape quality. Moreover, vines in this study were mechanically trimmed and harvested, which is becoming more common in the wine grape industry (Kaan Kurtural and Fidelibus, 2021). As a result, robust models which work reliably in vineyards managed mechanically will become more important in the future. Additionally, previous work has relied on hand-weighed yield data collected at the vine level, which is typically not implemented by commercial growers on a large scale. Instead, yield monitor data is more commonly used (Bramley and Hamilton, 2004), which allows for a larger dataset but also introduces additional sources of error into the yield estimates due to both the continuous nature of data collection as well as the influence of yield monitor geometry and mass flow on the measurement (Searcy et al., 1989).

5.4.4. Visualization and Saliency Mapping

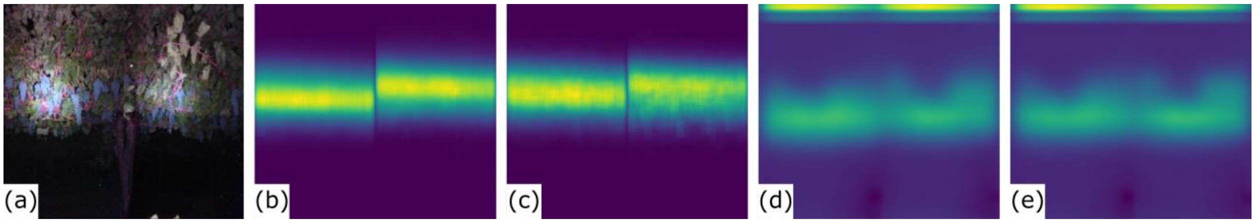


Figure 5.7. (a) Example input image to the CNN model. (b) Heatmap representing localization of predicted grape bunches using the object detection model across the train set and c) test set. (d) Grad-CAM heatmap representing areas averaged over all test set yield points with high influence on increasing predicted yield in the train set and e) test set.

Figure 5.7 contains an example input image to the CNN model, along with heatmaps representing the position of detected grapes on the vine in the training (Fig. 5.7b) test set (Fig 5.7c) as well as the average Grad-CAM response from the CNN network for the training and test sets (Fig 5.7d-e). In this visualization, although grape clusters are localized in the center of the image frame, the regions of the image with the strong contribution to predicted yield include the center of the frame as well as the top of the frame, potentially representing the density of the vine canopy at the top of the frame, where in some images, dark background is visible through the canopy. The fact that the CNN model was free to use these features of the input, whereas the object detection model was constrained to regions with grape bunches, may have contributed to the increased performance of both the CNN and transformer models over the object detection models. Alternatively, the attention on the top edge may also be a result of overfitting to the dataset by relating spurious image features with yield. This behavior was seen in both the training and test sets (Fig. 5.7d-e), suggesting that the attention was learned by the model, and not due to differences between the datasets.

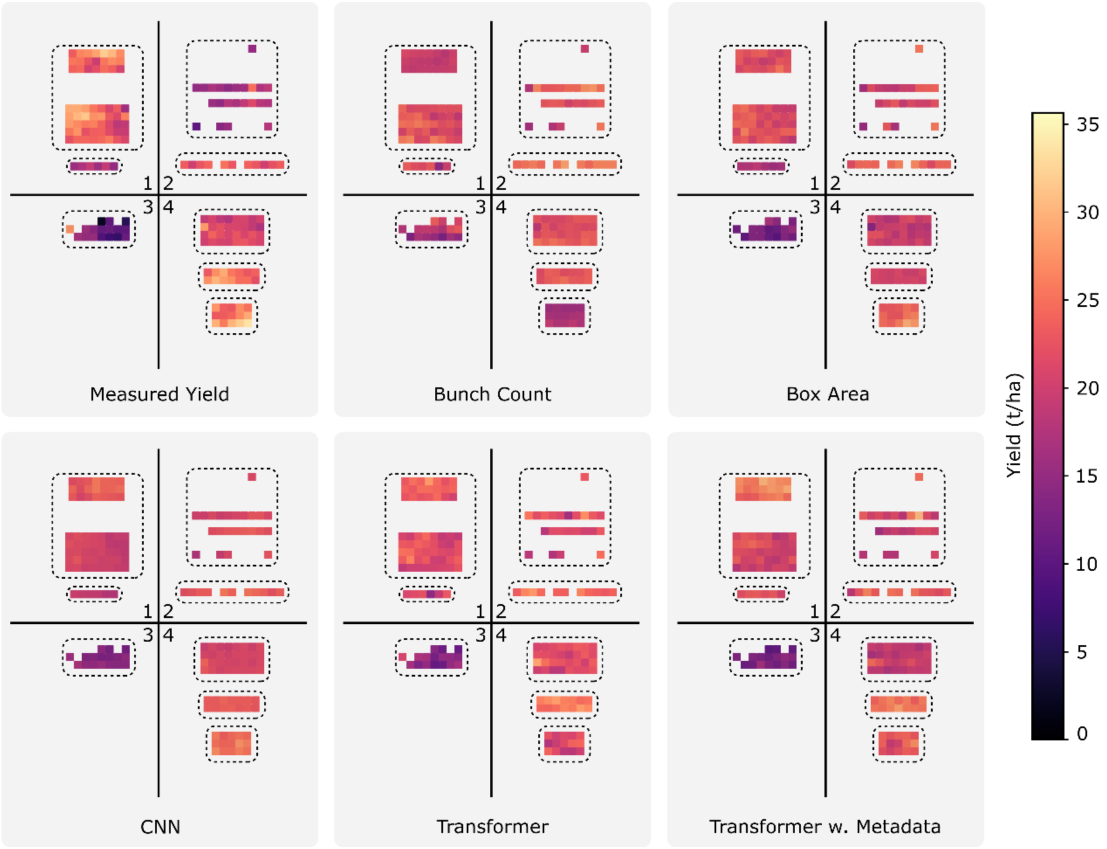


Figure 5.8: Yield maps from the test set (represented by dark points in Fig. 5.1). Maps were generated by aggregating data to 10m. Numbers in each pane represent block numbers, as in Fig. 5.1.

Yield maps generated from measured and predicted data demonstrate how the models tend to predict values close to the center of the distribution of values (Fig. 5.8). This is a visual representation of the range expressed results in Table 1, which demonstrate that no predicted dataset was able to achieve the same range of values as the measured data (< 100% range expressed). Nevertheless, visual trends from the measured plots can be seen in the predicted maps. Most notably, block 3 is represented on all maps with the lowest yield values. Likewise, the pattern shown in block 2 by both transformer models is similar to that of the measured data. Future work will involve collecting more data from more contiguous regions such that more detailed maps may be generated with the proximal imaging approach used in this work.

5.5. Conclusions

Three different models for prediction of grape yield from proximal imagery were trained based on image data collected from a vehicle-mounted sensing kit against ground truth yield data collected from a mechanical yield harvester. The object detection model, which is most similar to previous methods of yield estimation from proximal imagery, demonstrated poor performance when the size of grape bunches was not accounted for (MAPE of 29.4% when aggregated to 10 m blocks). Performance improved with use of grape bunch area as opposed to bunch count, with a MAPE of 18.5% at 10 m spatial aggregation.

Regression-based deep learning models trained end-to-end on yield prediction from input imagery demonstrated similar performance to the best object detection approach (grape bunch area). Regression CNN and transformer architectures were utilized, with an MAPE of up to 18% at 10 m spatial aggregation achieved with a transformer architecture after the addition of encoded positional metadata. Moreover, these regression architectures eliminate the need for hand-labeling images, removing a considerable bottleneck from the model development process.

While the performance on the test set in this work is lower than some previous studies, previous works have primarily been performed in vineyards trimmed by hand with vertical shoot positioned trellis configurations. This study represents an application of yield prediction in challenging conditions for image collection, with high occlusion, dense foliage, and a quadrilateral trellis configuration, common to the California Central Valley. Additionally, this study assessed performance on a holdout test set, which is more representative of unseen data.

Future work in this area will encompass yield forecasting from earlier in the season, at times prior to veraison, where accurate yield estimates are difficult to obtain but highly valued. Additionally, while the image data in this work were collected specifically for this study, the low-cost, vehicle-mounted sensing kit used for imaging may allow for automated data collection during routine management operations in the future, if the kit is mounted on existing equipment. Finally, due to the flexibility of the regression-based approach, data fusion techniques where proximal imagery is supplemented with remote sensing data represents a promising area for exploration in which coarse resolution satellite data may be used to improve performance of models trained on ground-based imagery.

CHAPTER 6. OVERALL CONCLUSIONS

This research consisted of acquisition and analysis of nondestructive image data in food and agricultural systems. Overall, the following conclusions were reached:

1. **Micro-CT images collected during time-series 3D imaging of apples during in vitro digestion demonstrated significant changes over time.** Images of apples in water and gastric juice both showed significant intensity differences from their initial intensity after 2.67 hours ($p < 0.05$). Additionally, the intensity of images of apples in gastric juice and water were significantly different from each other ($p < 0.05$) at all measured time points after 8 hours of incubation.
2. **Analysis of micro-CT images allowed for quantification of changes in food structure over time during in vitro digestion, which were similar to changes observed in destructive measurement of hardness.** Moisture uptake in apples during in vitro digestion, measured destructively, was similar between the two treatments tested: gastric juice and water. However, the destructively measured hardness of apples incubated in gastric juice decreased faster than apples in water. k (h^{-1}) was 0.15 ± 0.005 h^{-1} and 0.036 ± 0.01 h^{-1} for apples incubated in gastric juice and water, respectively (Eqn. 3.2). Correspondingly, the image-derived intensity of apples incubated in gastric juice changed at a faster rate than apples incubated in water.
3. **Custom convolutional neural networks trained in an end-to-end fashion were able to directly predict compression profiles from 3D image data of food during in vitro gastric digestion collected with micro-CT.** A well-known convolutional architecture, ResNet, was modified to accept 3D image data and output a vector of force values, representing the compression profile during mechanical testing. Moreover, the model was

trained to convergence with known compression curves in vector format, allowing for model prediction on unseen samples. Under cross-validation, the model trained on micro-CT images of apple tissue during in vitro digestion in gastric juice or water was able to predict compression curves for 3D images with a mean absolute percent error of 26.4% (39.7 and 13.0 % for gastric and water, respectively). The overall R^2 of the model predictions was 0.939, signifying good accordance with measured data.

4. **Compression curves predicted from micro-CT images of apples incubated in gastric juice and water can be analyzed to extract accurate time-dependent trends in tissue softening.** Extraction of peak force from compression curves assessed using cross-validation allowed for generation of time-dependent hardness profiles from the tested samples. These profiles could be fit to a Weibull model (Eqn. 3.2) with R^2 for the model fit of 0.82 and 0.72 for apples in gastric juice and water, respectively. Under cross-validation, the hardness profiles matched well with destructive data, with a mean absolute percent error of 23 and 13.4% for apples in gastric juice and water, respectively. However, when an out-of-distribution variability set was assessed, the performance was inconsistent, with a mean absolute percent error of 102.4 and 8.57% for apples in gastric juice and water, respectively.
5. **Large-scale grape yield-monitor observations were accurately predicted from RGB image data collected with consumer cameras in a mechanically managed commercial vineyard setting after aggregation to 10 m spatial bins.** Three deep learning architectures: object detection, convolutional neural network regression, and a transformer network were all able to predict grape yield from image data to within varying degrees of accuracy, with a mean absolute percent error of yield estimation of as

low as 18% when image and yield data were aggregated to 10 m regions. Notably, this result was achieved on a representative holdout set within a mechanically managed vineyard, which contrasts with previous studies in the field that have concentrated on correlation results in vineyards with sparse, hand-trimmed canopies.

6. **End-to-end modeling for regression tasks allowed for similar predictive performance of large-scale yield-monitor datasets collected in a commercial vineyard in comparison to object detection-based models without the need for hand-labeling.** Performance achieved by the end-to-end models in terms of mean absolute percent error (MAPE) was slightly better than performance achieved by analysis of object detection results, depending on the extent to which the yield data was spatially aggregated. At 10 m of spatial aggregation, the transformer model demonstrated slightly higher performance than the object detection model, with a 18% for the transformer model and 18.5% for the object detection model. These results for the end-to-end models were achieved without the need for hand-labeling, reducing the human labor required for incorporating data into yield estimation models.

CHAPTER 7. FUTURE WORK

Future work in this field will involve the collection and analysis of more data in more settings to determine the robustness of the approach described throughout this work, namely, nondestructive imaging for quantitative analysis using deep learning. In foods, micro-CT images of apples during in vitro digestion represented an effective combination of imaging modality, food system, and time-dependent process. However, additional imaging modalities such as color, hyperspectral, or magnetic resonance imaging represent promising areas for exploration with a wider array of food materials. Additionally, due to the nondestructive nature of the approach described in this work, processes such as food drying, freezing, and storage represent avenues for future exploration. In the agricultural domain, prediction of yield from images of grapevines collected shortly before harvest represented an important first step in determining the feasibility of the nondestructive approach. However, prediction of yield from earlier in the growing season would be more useful for growers, so management may be altered during the growing season, instead of between seasons. Additionally, specialty crops other than grapevines, such as almonds or strawberries, may benefit from a similar approach. Finally, the nondestructive and rapid nature of analysis may allow for utility in the field of crop phenotyping, where manual assessment of plant properties in the field represents a significant bottleneck in the development of new varieties of food crops.

REFERENCES

Aguilera, J.M., 2005. Why food microstructure? J. Food Eng. 67, 3–11. https://doi.org/10.1016/j.jfoodeng.2004.05.050

Aguilera, J.M., Lillford, P.J., 2008. Structure–Property Relationships in Foods, in: Food Materials Science. Springer New York, New York, NY, pp. 229–253. https://doi.org/10.1007/978-0-387-71947-4_12

Alam, T., Takhar, P.S., 2016. Microstructural Characterization of Fried Potato Disks Using X-Ray Micro Computed Tomography. J. Food Sci. 81, E651–E664. https://doi.org/10.1111/1750-3841.13219

Anastasiou, E., Balafoutis, A., Darra, N., Psiroukis, V., Biniari, A., Xanthopoulos, G., Fountas, S., 2018. Satellite and proximal sensing to estimate the yield and quality of table grapes. Agric. 8. https://doi.org/10.3390/agriculture8070094

AOAC, 2000. Method 934.06 Moisture in Dried Fruits. AOAC Int. 215, 1.

Aquino, A., Millan, B., Diago, M.P., Tardaguila, J., 2018. Automated early yield prediction in vineyards from on-the-go image acquisition. Comput. Electron. Agric. 144, 26–36. https://doi.org/10.1016/j.compag.2017.11.026

Arthur, D., Vassilvitskii, S., 2007. K-Means++: the Advantages of Careful Seeding. Proc. eighteenth Annu. ACM-SIAM Symp. Discret. algorithms 1027–1025. https://doi.org/10.1145/1283383.1283494

Baker, D.R., Mancini, L., Polacci, M., Higgins, M.D., Gualda, G.A.R., Hill, R.J., Rivers, M.L., 2012. An introduction to the application of X-ray microtomography to the three-dimensional study of igneous rocks. Lithos 148, 262–276. https://doi.org/10.1016/j.lithos.2012.06.008

Ballesteros, R., Intrigliolo, D.S., Ortega, J.F., Ramírez-Cuesta, J.M., Buesa, I., Moreno, M.A., 2020. Vineyard yield estimation by combining remote sensing, computer vision and artificial neural network techniques. Precis. Agric. 21, 1242–1262. https://doi.org/10.1007/s11119-020-09717-3

Bargoti, S., Underwood, J., 2015. Utilising Metadata to Aid Image Classification in Orchards. IEEE Int. Conf. Intell. Robot. Syst. (IROS), Work. Altern. Sens. Robot Percept. 1–3.

Bargoti, S., Underwood, J.P., 2017. Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards. J. F. Robot. 34, 1039–1060. https://doi.org/10.1002/rob.21699

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020. Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Barrett, J.F., Keat, N., 2004. Artifacts in CT: Recognition and Avoidance. RadioGraphics 24, 1679–1691. https://doi.org/10.1148/rg.246045065

Barron, J.T., 2019. A general and adaptive robust loss function. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019-June, 4326–4334. https://doi.org/10.1109/CVPR.2019.00446

Beaulieu, M., Turgeon, S.L., Doublier, J.L., 2001. Rheology, texture and microstructure of whey proteins/low methoxyl pectins mixed gels with added calcium. Int. Dairy J. 11, 961–967. https://doi.org/10.1016/S0958-6946(01)00127-3

Bernin, D., Steglich, T., Röding, M., Moldin, A., Topgaard, D., Langton, M., 2014. Multi-scale characterization of pasta during cooking using microscopy and real-time magnetic resonance imaging. Food Res. Int. 66, 132–139. https://doi.org/10.1016/j.foodres.2014.09.007

Bolya, D., Zhou, C., Xiao, F., Lee, Y.J., 2020. YOLACT++: Better Real-time Instance Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. X, 1–13. https://doi.org/10.1109/TPAMI.2020.3014297

Bolya, D., Zhou, C., Xiao, F., Lee, Y.J., 2019. YOLACT: Real-time instance segmentation. Proc. IEEE Int. Conf. Comput. Vis. 2019-Octob, 9156–9165. https://doi.org/10.1109/ICCV.2019.00925

Borlaug, N.E., 2002. Feeding a World of 10 Billion People : The Miracle Ahead. Vitr. Cell. Dev. Biol. - Plant 38, 221–228. https://doi.org/10.1079/1VP2001279

Bornhorst, G.M., Ferrua, M.J., Rutherfurd, S.M., Heldman, D.R., Singh, R.P., 2013. Rheological Properties and Textural Attributes of Cooked Brown and White Rice During Gastric Digestion in Vivo 137–150. https://doi.org/10.1007/s11483-013-9288-1

Bornhorst, G.M., Ferrua, M.J., Singh, R.P., 2015. A Proposed Food Breakdown Classification System to Predict Food Behavior during Gastric Digestion. J. Food Sci. 80, R924–R934. https://doi.org/10.1111/1750-3841.12846

Bornhorst, G.M., Gouseti, O., Wickham, M.S.J., Bakalis, S., 2016. Engineering Digestion: Multiscale Processes of Food Digestion. J. Food Sci. 81, R534–R543. https://doi.org/10.1111/1750-3841.13216

Bornhorst, G.M., Singh, P.R., 2014. Gastric digestion in vivo and in vitro: how the structural aspects of food influence the digestion process. Annu. Rev. Food Sci. Technol. 5, 111–32. https://doi.org/10.1146/annurev-food-030713-092346

Bornhorst, G.M., Singh, R.P., 2013. Kinetics of in Vitro Bread Bolus Digestion with Varying Oral and Gastric Digestion Parameters. Food Biophys. 8, 50–59. https://doi.org/10.1007/s11483-013-9283-6

Bourne, M., 2002. Food Texture and Viscosity: Concept and Measurement. Elsevier Science & Technology Books.

Bourne, M.C., 2002. Food Texture and Viscosity, in: Elsevier Science & Technology Books. pp. 305–309. https://doi.org/10.1016/S0924-6509(08)70056-4

Bramley, R.G.V., Hamilton, R.P., 2004. Understanding variability in winegrape production systems 1. Within vineyard variation in yield over several vintages. Aust. J. Grape Wine Res. 10, 32–45. https://doi.org/10.1111/j.1755-0238.2004.tb00006.x

Bultreys, T., Boone, M.A., Boone, M.N., Schryver, T. De, Masschaele, B., Hoorebeke, L. Van, Cnudde, V., 2016. Advances in Water Resources Fast laboratory-based micro-computed tomography for pore-scale research : Illustrative experiments and perspectives on the future. Adv. Water Resour. 95, 341–351. https://doi.org/10.1016/j.advwatres.2015.05.012

Capuano, E., Janssen, A.E.M., 2021. Food Matrix and Macronutrient Digestion. Annu. Rev. Food Sci. Technol. 12, 193–212. https://doi.org/10.1146/annurev-food-032519-051646

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-End Object Detection with Transformers. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 12346 LNCS, 213–229. https://doi.org/10.1007/978-3-030-58452-8_13

Carrillo, E., Matese, A., Rousseau, J., Tisseyre, B., 2016. Use of multi-spectral airborne imagery to improve yield sampling in viticulture. Precis. Agric. 17, 74–92. https://doi.org/10.1007/s11119-015-9407-8

Chen, J., Gaikwad, V., Holmes, M., Murray, B., Povey, M., Wang, Y., Zhang, Y., 2011. Development of a simple model device for in vitro gastric digestion investigation. Food Funct. 2, 174–182. https://doi.org/10.1039/c0fo00159g

Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking Atrous Convolution for Semantic Image Segmentation.

Chen, L., Opara, U.L., 2013. Texture measurement approaches in fresh and processed foods - A review. Food Res. Int. 51, 823–835. https://doi.org/10.1016/j.foodres.2013.01.046

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40, 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2015. Semantic Image Segmanetation with Deep Convolutional Nets and Fully Connected CRF, in: ICLR 2015.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding.

Cunha, M., Marçal, A.R.S., Silva, L., 2010. Very early prediction of wine yield based on satellite data from vegetation. Int. J. Remote Sens. 31, 3125–3142. https://doi.org/10.1080/01431160903154382

Dalmau, M.E., Bornhorst, G.M., Eim, V., Rosselló, C., Simal, S., 2017. Effects of freezing, freeze drying and convective drying on in vitro gastric digestion of apples. Food Chem. 215, 7–16. https://doi.org/10.1016/j.foodchem.2016.07.134

De La Fuente, M., Linares, R., Baeza, P., Miranda, C., Lissarrague, J.R., 2015. Comparison of different methods of grapevine yield prediction in the time window between fruitset and veraison. J. Int. des Sci. la Vigne du Vin 49, 27–35. https://doi.org/10.20870/oeno-one.2015.49.1.96

Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-fei, L., 2009. ImageNet : A Large-Scale Hierarchical Image Database 248–255.

Dhillon, A., Verma, G.K., 2020. Convolutional neural network: a review of models, methodologies and applications to object detection. Prog. Artif. Intell. 9, 85–112. https://doi.org/10.1007/s13748-019-00203-0

Dhondt, S., Vanhaeren, H., Van Loo, D., Cnudde, V., Inzé, D., 2010. Plant structure visualization by high-resolution X-ray computed tomography. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2010.05.002

Di Gennaro, S.F., Toscano, P., Cinat, P., Berton, A., Matese, A., 2019. A low-cost and unsupervised image recognition methodology for yield estimation in a vineyard. Front. Plant Sci. 10, 1–13. https://doi.org/10.3389/fpls.2019.00559

Diago, M.P., Correa, C., Millán, B., Barreiro, P., Valero, C., Tardaguila, J., 2012. Grapevine yield and leaf area estimation using supervised classification methodology on RGB images taken under field conditions. Sensors (Switzerland) 12, 16988–17006. https://doi.org/10.3390/s121216988

Diels, E., van Dael, M., Keresztes, J., Vanmaercke, S., Verboven, P., Nicolai, B., Saeys, W., Ramon, H., Smeets, B., 2017. Assessment of bruise volumes in apples using X-ray computed tomography. Postharvest Biol. Technol. 128, 24–32. https://doi.org/10.1016/j.postharvbio.2017.01.013

Donis-González, I.R., Guyer, D.E., Fulbright, D.W., Pease, A., 2014. Postharvest noninvasive assessment of fresh chestnut (Castanea spp.) internal decay using computer tomography images. Postharvest Biol. Technol. 94, 14–25. https://doi.org/10.1016/j.postharvbio.2014.02.016

Donis-González, I.R., Guyer, D.E., Leiva-Valenzuela, G.A., Burns, J., 2013. Assessment of chestnut (Castanea spp.) slice quality using color images. J. Food Eng. 115, 407–414. https://doi.org/10.1016/j.jfoodeng.2012.09.017

Donis-González, I.R., Guyer, D.E., Pease, A., 2016. Postharvest noninvasive assessment of undesirable fibrous tissue in fresh processing carrots using computer tomography images. Postharvest Biol. Technol. 190, 154–166. https://doi.org/10.1016/j.postharvbio.2016.06.024

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.

Drechsler, K.C., Bornhorst, G.M., 2018. Modeling the softening of carbohydrate-based foods during simulated gastric digestion. J. Food Eng. 222, 38–48. https://doi.org/10.1016/j.jfoodeng.2017.11.007

Drechsler, K.C., Ferrua, M.J., 2015. Modelling the breakdown mechanics of solid foods during gastric digestion. Food Res. Int. 88, 181–190. https://doi.org/10.1016/j.foodres.2016.02.019

Duda, R.O., Hart, P.E., Stork, D.G., 2000. Pattern Classification. New York John Wiley, Sect. https://doi.org/10.1007/BF01237942

Dye, L., Blundell, J., 2002. Functional foods: psychological and behavioural functions. Br. J. Nutr. 88, S187. https://doi.org/10.1079/BJN2002684

E.J. Gallo Winery, 2020. Personal Communication.

Earles, J.M., Knipfer, T., Tixier, A., Orozco, J., Reyes, C., Zwieniecki, M.A., Brodersen, C.R., McElrone, A.J., 2018. In vivo quantification of plant starch reserves at micrometer resolution using X-ray microCT imaging and machine learning. New Phytol. 218, 1260–1269. https://doi.org/10.1111/nph.15068

Eboibi, O., Uguru, H., 2017. Storage conditions effect on physical, mechanical and textural properties of intact cucumber (cv Nandini) fruit 0869, 48–56.

Ege, T., Yanai, K., 2017. Simultaneous estimation of food categories and calories with multi-task CNN. Proc. 15th IAPR Int. Conf. Mach. Vis. Appl. MVA 2017 198–201. https://doi.org/10.23919/MVA.2017.7986835

Eggert, A., Müller, M., Nachtrab, F., Dombrowski, J., Rack, A., Zabler, S., 2014. High-speed in-situ tomography of liquid protein foams. Int. J. Mater. Res. 105, 632–639. https://doi.org/10.3139/146.111057

Ellis, P.R., Kendall, C.W.C., Ren, Y., Parker, C., Pacy, J.F., Waldron, K.W., Jenkins, D.J.A., 2004. Role of cell walls in the bioaccessibility of lipids in almond seeds. Am. J. Clin. Nutr. 80, 604–613. https://doi.org/80/3/604 [pii]

Fei, Z., Olenskyj, A.G., Bailey, B.N., Earles, M., 2021. Enlisting 3D Crop Models and GANs for More Data Efficient and Generalizable Fruit Detection 1269–1277.

Feldkamp, L.A.A., Davis, L.C.C., Kress, J.W.W., 1984. Practical cone-beam algorithm. J. Opt. Soc. Am. 1, 612–619. https://doi.org/10.1364/JOSAA.1.000612

Fortuny, R.C.S., Lluch, M.A., Quiles, A., Miguel, N.G., Belloso, O.M., 2003. Evaluation of textural properties and microstructure during storage of minimally processed apples. J. Food Sci. 68, 312–317. https://doi.org/10.1111/j.1365-2621.2003.tb14158.x

Frank Rosenblatt, 1958. The Perceptron: a Probabilistic Model for Information Storage and Organization in the Brain. Psychol. Rev. 65, 386–408.

Fu, H., Gong, M., Wang, C., Batmanghelich, K., Tao, D., 2018. Deep Ordinal Regression Network for Monocular Depth Estimation Huan, in: CVPR 2018. pp. 2002–2011.

Fukushima, K., 1980. Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position 36, 193–202.

Gené-Mola, J., Vilaplana, V., Rosell-Polo, J.R., Morros, J.-R., Ruiz-Hidalgo, J., Gregorio, E., 2019. Multi-modal deep learning for Fuji apple detection using RGB-D cameras and their radiometric capabilities. Comput. Electron. Agric. 162, 689–698.

Ghaitaranpour, A., Rastegar, A., Tabatabaei Yazdi, F., Mohebbi, M., Alizadeh Behbahani, B., 2017. Application of Digital Image Processing in Monitoring some Physical Properties of Tarkhineh during Drying. J. Food Process. Preserv. 41, 1–9. https://doi.org/10.1111/jfpp.12861

Girshick, R., 2015. Fast R-CNN. Proc. IEEE Int. Conf. Comput. Vis. 2015 Inter, 1440–1448. https://doi.org/10.1109/ICCV.2015.169

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 580–587. https://doi.org/10.1109/CVPR.2014.81

Gongal, A., Amatya, S., Karkee, M., Zhang, Q., Lewis, K., 2015. Sensors and systems for fruit detection and localization: A review. Comput. Electron. Agric. 116, 8–19. https://doi.org/10.1016/j.compag.2015.05.021

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. NIPS 2014.

Grassby, T., Picout, D.R., Mandalari, G., Faulks, R.M., Kendall, C.W.C.C., Rich, G.T., Wickham, M.S.J.J., Lapsley, K., Ellis, P.R., 2014. Modelling of nutrient bioaccessibility in almond seeds based on the fracture properties of their cell walls. Food Funct. 5, 3096–3106. https://doi.org/10.1039/c4fo00659c

Groß, D., Zick, K., Guthausen, G., 2017. Recent MRI and Diffusion Studies of Food Structures, 1st ed, Annual Reports on NMR Spectroscopy. Elsevier Ltd. https://doi.org/10.1016/bs.arnmr.2016.09.001

Grundy, M.M.L., Carriere, F., Mackie, A.R., Gray, D.A., Butterworth, P.J., Ellis, P.R., 2016a. The role of plant cell wall encapsulation and porosity in regulating lipolysis during the digestion of almond seeds. Food Funct. 7, 69–78. https://doi.org/10.1039/c5fo00758e

Grundy, M.M.L., Grassby, T., Mandalari, G., Waldron, K.W., Butterworth, P.J., Berry, S.E.E., Ellis, P.R., 2015. Effect of mastication on lipid bioaccessibility of almonds in a randomized human study and its implications for digestion kinetics, metabolizable energy, and postprandial lipemia. Am. J. Clin. Nutr. 101, 25–33. https://doi.org/10.3945/ajcn.114.088328

Grundy, M.M.L., Lapsley, K., Ellis, P.R., 2016b. A review of the impact of processing on nutrient bioaccessibility and digestion of almonds. Int. J. Food Sci. Technol. 51, 1937–1946. https://doi.org/10.1111/ijfs.13192

Gulati, T., Datta, A.K., Doona, C.J., Ruan, R.R., Feeherry, F.E., 2015. Modeling moisture migration in a multi-domain food system: Application to storage of a sandwich system. Food Res. Int. 76, 427–438. https://doi.org/10.1016/j.foodres.2015.06.022

Guo, S., Bocklitz, T., Neugebauer, U., Popp, J., 2017. Common mistakes in cross-validating classification models. Anal. Methods 9, 4410–4417. https://doi.org/10.1039/c7ay01363a

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M.S., 2016. Deep learning for visual understanding: A review. Neurocomputing 187, 27–48. https://doi.org/10.1016/j.neucom.2015.09.116

Hafiz, A.M., Bhat, G.M., 2020. A survey on instance segmentation: state of the art. Int. J. Multimed. Inf. Retr. 9, 171–189. https://doi.org/10.1007/s13735-020-00195-x

Halford, J.C.G., Harrold, J.A., 2012. Satiety-enhancing products for appetite control: science and regulation of functional foods for weight management. Proc. Nutr. Soc. 71, 350–362. https://doi.org/10.1017/S0029665112000134

Hamidinekoo, A., Denton, E., Rampun, A., Honnor, K., Zwiggelaar, R., 2018. Deep learning in mammography and breast histology, an overview and future trends. Med. Image Anal. 47, 45–67. https://doi.org/10.1016/j.media.2018.03.006

Häni, N., Roy, P., 2019. A comparative study of fruit detection and counting methods for yield mapping in apple orchards. https://doi.org/10.1002/rob.21902

Hayman, P., Longbottom, M., McCarthy, M., Thomas, D., 2012. Managing vines during heatwaves. Wine Aust. Aust. Wine Factsheet 1–8.

He, K., Gkioxari, G., Dollar, P., Girshick, R., 2017. Mask R-CNN. Proc. IEEE Int. Conf. Comput. Vis. 2017-Octob, 2980–2988. https://doi.org/10.1109/TPAMI.2018.2844175

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hellman, E., 2004. How to Judge Grape Ripeness Before Harvest. Southwest Reg. Vine Wine Conf.

Herbstreith & Fox, 2000. High Quality resulting from Product Integrated Environment Protection - PIUS, Fruit Processing. https://doi.org/10.1038/266006a0

Herremans, E., Melado-herreros, A., Defraeye, T., Verlinden, B., Bongaers, E., Estrade, P., Wevers, M., Barreiro, P., Nicolaï, B.M., 2014a. Comparison of X-ray CT and MRI of watercore disorder of different apple cultivars. Postharvest Biol. Technol. 87, 42–50.

Herremans, E., Verboven, P., Bongaers, E., Estrade, P., Verlinden, B.E., Wevers, M., Hertog, M.L.A.T.M., Nicolai, B.M., 2013. Characterisation of "Braeburn" browning disorder by means of X-ray micro-CT. Postharvest Biol. Technol. 75, 114–124. https://doi.org/10.1016/j.postharvbio.2012.08.008

Herremans, E., Verboven, P., Defraeye, T., Rogge, S., Ho, Q.T., Hertog, M.L.A.T.M., Verlinden, B.E., Bongaers, E., Wevers, M., Nicolai, B.M., 2014b. X-ray CT for quantitative food microstructure engineering: The apple case. Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms 324, 88–94. https://doi.org/10.1016/j.nimb.2013.07.035

Herrero-Huerta, M., González-Aguilera, D., Rodriguez-Gonzalvez, P., Hernández-López, D., 2015. Vineyard yield estimation by automatic 3D bunch modelling in field conditions. Comput. Electron. Agric. 110, 17–26. https://doi.org/10.1016/j.compag.2014.10.003

Herriott, C., Spear, A.D., 2020. Predicting microstructure-dependent mechanical properties in additively manufactured metals with machine- and deep-learning methods. Comput. Mater. Sci. 175, 109599. https://doi.org/10.1016/j.commatsci.2020.109599

Heyes, J.A., Clark, C.J., 2003. Magnetic resonance imaging of water movement through asparagus. Funct. Plant Biol. 30, 1089–1095. https://doi.org/10.1071/FP03096

Ho, Q.T., Carmeliet, J., Datta, A.K., Defraeye, T., Delele, M.A., Herremans, E., Opara, L., Ramon, H., Tijskens, E., Van Der Sman, R., Van Liedekerke, P., Verboven, P., Nicolaï, B.M., 2013. Multiscale modeling in food engineering. J. Food Eng. 114, 279–291. https://doi.org/10.1016/j.jfoodeng.2012.08.019

Ho, Q.T., Verboven, P., Verlinden, B.E., Herremans, E., Wevers, M., Carmeliet, J., Nicolai, B.M., 2011. A three-dimensional multiscale model for gas exchange in fruit. Plant Physiol 155, 1158–1168. https://doi.org/10.1104/pp.110.169391

Hopkinson, I., Jones, R.A.L., Black, S., Lane, D.M., McDonald, P.J., 1997. Fickian and case II diffusion of water into amylose: A stray field NMR study. Carbohydr. Polym. 34, 39–47. https://doi.org/10.1016/S0144-8617(97)00106-9

Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Le, Q., Adam, H., 2019. Searching for MobileNetV3. Proc. IEEE Int. Conf. Comput. Vis. 2019-Octob, 1314–1324. https://doi.org/10.1109/ICCV.2019.00140

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.

Hur, S.J., Decker, E.A., McClements, D.J., 2009. Influence of initial emulsifier type on microstructural changes occurring in emulsified lipids during in vitro digestion. Food Chem. 114, 253–262. https://doi.org/10.1016/j.foodchem.2008.09.069

Hur, S.J., Lim, B.O., Decker, E.A., McClements, D.J., 2011. In vitro human digestion models for food applications. Food Chem. 125, 1–12. https://doi.org/10.1016/j.foodchem.2010.08.036

Hutchings, S.C., Foster, K.D., Bronlund, J.E., Lentle, R.G., Jones, J.R., Morgenstern, M.P., 2011. Mastication of heterogeneous foods: Peanuts inside two different food matrices. Food Qual. Prefer. 22, 332–339. https://doi.org/10.1016/j.foodqual.2010.12.004

Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-January, 5967–5976. https://doi.org/10.1109/CVPR.2017.632

Jocher, G., Kwon, Y., guigarfr, perry0418, Veitch-Michaelis, J., Ttayu, Suess, D., Baltacı, F., Bianconi, G., IlyaOvodov, Marc, e96031413, Lee, C., Kendall, D., Falak, Reveriano, F., FuLin, GoogleWiki, Nataprawira, J., Hu, J., LinCoce, LukeAI, NanoCode012, NirZarrabi, Reda, O., Skalski, P., SergioSanchezMontesUAM, Song, S., Havlik, T., Shead, T.M., 2021a. ultralytics/yolov3: v9.5.0 - YOLOv5 v5.0 release compatibility update for YOLOv3. https://doi.org/10.5281/zenodo.4681234

Jocher, G., Stoken, A., Borovec, J., NanoCode012, ChristopherSTAN, Changyu, L., Laughing, tkianai, yxNONG, Hogan, A., lorenzomammana, AlexWang1900, Chaurasia, A., Diaconu, L., Marc, wanghaoyang0106, ml5ah, Doug, Durgesh, Ingham, F., Frederik, Guilhen, Colmagro, A., Ye, H., Jacobsolawetz, Poznanski, J., Fang, J., Kim, J., Doan, K., 于力军 L.Y., 2021b. ultralytics/yolov5: v4.0 - nn.SiLU() activations, Weights &amp; Biases logging, PyTorch Hub integration. https://doi.org/10.5281/ZENODO.4418161

Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sens. Environ. 141, 116–128. https://doi.org/10.1016/j.rse.2013.10.027

Kaan Kurtural, S., Fidelibus, M.W., 2021. Mechanization of Pruning, Canopy Management, and Harvest in Winegrape Vineyards. Catal. Discov. into Pract. 5, 29–44. https://doi.org/10.5344/catalyst.2021.20011

Kakani, V., Nguyen, V.H., Kumar, B.P., Kim, H., Pasupuleti, V.R., 2020. A critical review on computer vision and artificial intelligence in food industry. J. Agric. Food Res. 2, 100033. https://doi.org/10.1016/j.jafr.2020.100033

Kaláb, M., Allan-Wojtas, P., Miller, S.S., 1995. Microscopy and other imaging techniques in food structure analysis. Trends Food Sci. Technol. 6, 177–186. https://doi.org/10.1016/S0924-2244(00)89052-4

Kalender, W.A., 2006. X-ray computed tomography. Phys. Med. Biol. 51, R29–R43. https://doi.org/10.1088/0031-9155/51/13/R03

Kamilaris, A., Kartakoullis, A., Prenafeta-boldú, F.X., 2017. A review on the practice of big data analysis in agriculture. Comput. Electron. Agric. 143, 23–37. https://doi.org/10.1016/j.compag.2017.09.037

Kamilaris, A., Prenafeta-Boldú, F.X., 2018. A review of the use of convolutional neural networks in agriculture. J. Agric. Sci. 156, 312–322. https://doi.org/10.1017/S0021859618000436

Karras, T., Laine, S., Aila, T., 2020. A Style-Based Generator Architecture for Generative Adversarial Networks. IEEE Trans. Pattern Anal. Mach. Intell. 43, 4217–4228. https://doi.org/10.1109/tpami.2020.2970919

Kaur, P., Sikka, K., Wang, W., Belongie, S., Divakaran, A., 2019. FoodX-251: A Dataset for Fine-grained Food Classification 2–6.

Kazmierski, M., Glemas, P., Rousseau, J., Tisseyre, B., 2011. Temporal stability of within-field patterns of ndvi in non irrigated mediterranean vineyards. J. Int. des Sci. la Vigne du Vin 45, 61–73. https://doi.org/10.20870/oeno-one.2011.45.2.1488

Khaliq, A., Comba, L., Biglia, A., Ricauda Aimonino, D., Chiaberge, M., Gay, P., 2019. Comparison of satellite and UAV-based multispectral imagery for vineyard variability assessment. Remote Sens. 11. https://doi.org/10.3390/rs11040436

Khan, A.A., Vincent, J.F. V., 1993. Compressive Stiffness and Fracture Properties of Apple and Potato Parenchyma. J. Texture Stud. 24, 423–435. https://doi.org/10.1111/j.1745-4603.1993.tb00052.x

Kim, S., Schatzki, T.F., 2000. Apple watercore sorting system using x-ray imagery: I. Algorithm development. Trans. Am. Soc. Agric. Eng. 43, 1695–1702. https://doi.org/10.13031/2013.3070

Kondjoyan, A., Daudin, J.D., Santé-Lhoutellier, V., 2015. Modelling of pepsin digestibility of myofibrillar proteins and of variations due to heating. Food Chem. 172, 265–271. https://doi.org/10.1016/j.foodchem.2014.08.110

Kong, F., Singh, R.P., 2011. Solid Loss of Carrots During Simulated Gastric Digestion. Food Biophys. 6, 84–93. https://doi.org/10.1007/s11483-010-9178-8

Kong, F., Singh, R.P., 2009a. Digestion of raw and roasted almonds in simulated gastric environment. Food Biophys. 4, 365–377. https://doi.org/10.1007/s11483-009-9135-6

Kong, F., Singh, R.P., 2009b. Modes of disintegration of solid foods in simulated gastric environment. Food Biophys. 4, 180–190. https://doi.org/10.1007/s11483-009-9116-9

Kong, F., Singh, R.P., 2008. A model stomach system to investigate disintegration kinetics of solid foods during gastric digestion. J. Food Sci. 73, 202–210. https://doi.org/10.1111/j.1750-3841.2008.00745.x

Krizhevsky, A., 2009. Learning Multiple Layers of Features from Tiny Images.

Krizhevsky, A., Hinton, G.E., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, NIPS'12. Curran Associates Inc., Red Hook, NY, USA, pp. 1097–1105.

Kwon, O., Park, T., 2017. Applications of Smartphone Cameras in Agriculture , Environment , and Food : A review 42, 330–338.

Lähteenmäki, L., 2013. Claiming health in food products. Food Qual. Prefer. 27, 196–201. https://doi.org/10.1016/j.foodqual.2012.03.006

Lammertyn, J., Dresselaers, T., Van Hecke, P., Jancsók, P., Wevers, M., Nicolaï, B.M., 2003. MRI and X-ray CT study of spatial distribution of core breakdown in "Conference" pears. Magn. Reson. Imaging 21, 805–815. https://doi.org/10.1016/S0730-725X(03)00105-X

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. https://doi.org/10.1038/nature14539

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Comput. 1, 541–551. https://doi.org/10.1162/neco.1989.1.4.541

LeCunn, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gardient-Based Learning Applied to Document Recognition. Proced. IEEE.

Li, H.T., Chen, S.Q., Bui, A.T., Xu, B., Dhital, S., 2021. Natural 'capsule' in food plants: Cell wall porosity controls starch digestion and fermentation. Food Hydrocoll. 117, 106657. https://doi.org/10.1016/j.foodhyd.2021.106657

Li, X., Liu, Z., Cui, S., Luo, C., Li, C., Zhuang, Z., 2019. Predicting the effective mechanical property of heterogeneous materials by image based modeling and deep learning. Comput. Methods Appl. Mech. Eng. 347, 735–753. https://doi.org/10.1016/j.cma.2019.01.005

Li, Y., Chen, L., 2014. Big Biological Data : Challenges and Opportunities. Genomics. Proteomics Bioinformatics 12, 187–189. https://doi.org/10.1016/j.gpb.2014.10.001

Liang, L., Liu, M., Sun, W., 2017. A deep learning approach to estimate chemically-treated collagenous tissue nonlinear anisotropic stress-strain responses from microscopy images. Acta Biomater. 63, 227–235. https://doi.org/10.1016/j.actbio.2017.09.025

Liang, L., Wu, X., Zhao, T., Zhao, J., Li, F., Zou, Y., Mao, G., Yang, L., 2012. In vitro bioaccessibility and antioxidant activity of anthocyanins from mulberry (Morus atropurpurea Roxb.) following simulated gastro-intestinal digestion. Food Res. Int. 46, 76–82. https://doi.org/10.1016/j.foodres.2011.11.024

Liang, Y., Li, J., 2017. Computer vision-based food calorie estimation: dataset, method, and experiment.

Lim, K.S., Barigou, M., 2004. X-ray micro-computed tomography of cellular food products. Food Res. Int. 37, 1001–1012. https://doi.org/10.1016/j.foodres.2004.06.010

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. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 8693 LNCS, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., Ma, Y., 2016. DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment, in: Chang, C.K., Chiari, L., Cao, Y., Jin, H., Mokhtari, M., Aloulou, H. (Eds.), Inclusive Smart Cities and Digital Health. Springer International Publishing, Cham, pp. 37–48.

Liu, F., Feng, S., Guo, Y., Li, Z., Chen, L., Handa, A., Zhang, Y., 2021. The rheological characteristics of soy protein isolate-glucose conjugate gel during simulated gastrointestinal digestion. Food Struct. 29, 100210. https://doi.org/10.1016/j.foostr.2021.100210

Liu, S., Cossell, S., Tang, J., Dunn, G., Whitty, M., 2017. A computer vision system for early stage grape yield estimation based on shoot detection. Comput. Electron. Agric. 137, 88–101. https://doi.org/10.1016/j.compag.2017.03.013

Liu, S., Qi, L., Qin, H., Shi, J., Jia, J., 2018. Path Aggregation Network for Instance Segmentation. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 8759–8768. https://doi.org/10.1109/CVPR.2018.00913

Liu, S., Zeng, X., Whitty, M., 2020. A vision-based robust grape berry counting algorithm for fast calibration-free bunch weight estimation in the field. Comput. Electron. Agric. 173, 105360. https://doi.org/10.1016/j.compag.2020.105360

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016. SSD: Single shot multibox detector. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 9905 LNCS, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Llull, P., Simal, S., Femenia, A., Benedito, J., Rosselló, C., 2002. The use of ultrasound velocity measurement to evaluate the textural properties of sobrassada from Mallorca. J. Food Eng. 52, 323–330. https://doi.org/10.1016/S0260-8774(01)00122-4

Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation, in: CVPR 2015. https://doi.org/10.5244/C.30.124

Luo, Q., Boom, R.M., Janssen, A.E.M., 2015. Digestion of protein and protein gels in simulated gastric environment. LWT - Food Sci. Technol. 63, 161–168. https://doi.org/10.1016/j.lwt.2015.03.087

Mandalari, G., Faulks, R.M., Rich, G.T., Lo Turco, V., Picout, D.R., Lo Curto, R.B., Bisignano, G., Dugo, P., Dugo, G., Waldron, K.W., Ellis, P.R., Wickham, M.S.J.J., 2008. Release of protein, lipid, and vitamin E from almond seeds during digestion. J. Agric. Food Chem. 56, 3409–3416. https://doi.org/10.1021/jf073393v

Matese, A., Di Gennaro, S.F., 2021. Beyond the traditional NDVI index as a key factor to mainstream the use of UAV in precision viticulture. Sci. Rep. 11, 1–13. https://doi.org/10.1038/s41598-021-81652-3

Mattila-Sandholm, T., Myllärinen, P., Crittenden, R., Mogensen, G., Fondén, R., Saarela, M., 2002. Technological challenges for future Probiotic foods. Int. Dairy J. 12, 173–182. https://doi.org/10.1016/S0958-6946(01)00099-1

McClements, D.J., Decker, E.A., Park, Y., Weiss, J., 2009. Structural Design Principles for Delivery of Bioactive Components in Nutraceuticals and Functional Foods, Critical Reviews in Food Science and Nutrition. https://doi.org/10.1080/10408390902841529

Mebatsion, H.K., Verboven, P., Ho, Q.T., Verlinden, B.E., Nicolaï, B.M., 2008. Modelling fruit (micro)structures, why and how? Trends Food Sci. Technol. 19, 59–66.

https://doi.org/10.1016/j.tifs.2007.10.003

Mendoza, F., Verboven, P., Mebatsion, H.K., Kerckhofs, G., Wevers, M., Nicolaï, B., 2007. Three-dimensional pore space quantification of apple tissue using X-ray computed microtomography. Planta 226, 559–570. https://doi.org/10.1007/s00425-007-0504-4

Mennah-Govela, Y.A., Bornhorst, G.M., 2016a. Mass transport processes in orange-fleshed sweet potatoes leading to structural changes during in vitro gastric digestion. J. Food Eng. 191, 48–57. https://doi.org/10.1016/j.jfoodeng.2016.07.004

Mennah-Govela, Y.A., Bornhorst, G.M., 2016b. Acid and moisture uptake in steamed and boiled sweet potatoes and associated structural changes during in vitro gastric digestion. Food Res. Int. 88, 247–255. https://doi.org/10.1016/j.foodres.2015.12.012

Mery, D., Chanona-Perez, J.J., Soto, A., Aguilera, J.M., Cipriano, A., Velez-Rivera, N., Arzate-Vazquez, I., Gutierrez-Lopez, G.F., 2010. Quality classification of corn tortillas using computer vision. J. Food Eng. 101, 357–364. https://doi.org/10.1016/j.jfoodeng.2010.07.018

Mezzenga, R., Schurtenberger, P., Burbidge, A., Michel, M., 2005. Understanding foods as soft materials. Nat. Mater. 4, 729–740. https://doi.org/10.1038/nmat1496

Michel, M., Sagalowicz, L., 2008. Probing food structure. Food Mater. Sci. Princ. Pract. 203–226. https://doi.org/10.1007/978-0-387-71947-4_11

Milella, A., Marani, R., Petitti, A., Reina, G., 2019. In-field high throughput grapevine phenotyping with a consumer-grade depth camera. Comput. Electron. Agric. 156, 293–306. https://doi.org/10.1016/j.compag.2018.11.026

Millan, B., Velasco-Forero, S., Aquino, A., Tardaguila, J., 2018. On-the-go grapevine yield estimation using image analysis and boolean model. J. Sensors 2018. https://doi.org/10.1155/2018/9634752

Minaee, S., Boykov, Y.Y., Porikli, F., Plaza, A.J., Kehtarnavaz, N., Terzopoulos, D., 2021. Image Segmentation Using Deep Learning: A Survey. IEEE Trans. Pattern Anal. Mach. Intell. 8828, 1–20. https://doi.org/10.1109/TPAMI.2021.3059968

Minekus, M., Alminger, M., Alvito, P., Ballance, S., Bohn, T., Bourlieu, C., Carrì, F., Boutrou, R., Corredig, F.M., Dupont, D., Dufour, F.C., Egger, L., Golding, M., Karakaya, L.S., Kirkhus, B., Le Feunteun, S., Lesmes, U., Macierzanka, A., Mackie, A., Marze, S., Mcclements, D.J., Enard, O., Recio, I., Santos, C.N., Singh, R.P., Vegarud, G.E., Wickham, M.S.J., Weitschies, W., Brodkorb, A., 2014. A standardised static in vitro digestion method suitable for food – an international consensus. Food Funct. Food Funct 5, 1113–1124. https://doi.org/10.1039/c3fo60702j

Miri, T., Bakalis, S., Bhima, S.D., Fryer, P.J., 2006. Use of X-ray Micro-CT to characterize structure phenomena during frying., in: International Union of Food Science and Technology. pp. 735–747. https://doi.org/10.1051/IUFoST:20060023

Morell, P., Fiszman, S., Llorca, E., Hernando, I., 2017. Designing added-protein yogurts: Relationship between in vitro digestion behavior and structure. Food Hydrocoll. 72, 27–34. https://doi.org/10.1016/j.foodhyd.2017.05.026

Mu, Y., Chen, T.S., Ninomiya, S., Guo, W., 2020. Intact detection of highly occluded immature tomatoes on plants using deep learning techniques. Sensors (Switzerland) 20, 1–16. https://doi.org/10.3390/s20102984

Nadia, J., Olenskyj, A.G., Stroebinger, N., Hodgkinson, S.M., Estevez, T.G., Subramanian, P., Singh, H., Paul, R., Bornhorst, G.M., 2021. Tracking physical breakdown of rice- and wheat- based foods with varying structures during gastric digestion and its influence on gastric emptying in a growing pig model. Food Funct. 12, 4349–4372. https://doi.org/10.1039/d0fo02917c

Nair, V., Hinton, G.E., 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. ICML.

Nasiri, A., Taheri-Garavand, A., Zhang, Y.D., 2019. Image-based deep learning automated sorting of date fruit. Postharvest Biol. Technol. 153, 133–141. https://doi.org/10.1016/j.postharvbio.2019.04.003

Nazarian, A., Snyder, B.D., Zurakowski, D., Müller, R., 2008. Quantitative micro-computed tomography: A non-invasive method to assess equivalent bone mineral density. Bone 43, 302–311. https://doi.org/10.1016/j.bone.2008.04.009

Nielsen, S.S., 2010. Food Analysis, Food Analysis. https://doi.org/10.1007/978-1-4419-1478-1

Nixon, M.S., Aguado, A.S., 2002. Feature Extraction and Image Processing, Feature Extraction & Image Processing for Computer Vision. https://doi.org/10.1016/B978-0-12-396549-3.00001-X

Norton, J.E., Wallis, G.A., Spyropoulos, F., Lillford, P.J., Norton, I.T., 2014. Designing food structures for nutrition and health benefits. Ann Rev Food Sci Technol 5, 177–95. https://doi.org/10.1146/annurev-food-030713-092315

Nouri, M., Nasehi, B., Goudarzi, M., Abdanan Mehdizadeh, S., 2018. Non-destructive Evaluation of Bread Staling Using Gray Level Co-occurrence Matrices. Food Anal. Methods 11, 3391–3395. https://doi.org/10.1007/s12161-018-1319-6

Novotny, J.A., Gebauer, S.K., Baer, D.J., 2012. Discrepancy between the Atwater factor predicted and empirically measured energy values of almonds in human diets. Am. J. Clin. Nutr. 96, 296–301. https://doi.org/10.3945/ajcn.112.035782

Nuske, S., Gupta, K., Narasimhan, S., Singh, S., 2014a. Modeling and Calibrating Visual Yield Estimates in Vineyards. Springer Tracts Adv. Robot. 92, 343–356. https://doi.org/10.1007/978-3-642-40686-7

Nuske, S., Wilshusen, K., Achar, S., Yoder, L., Narasimhan, S., Singh, S., 2014b. Automated Visual Yield Estimation in Vineyards. J. F. Robot. 31, 837–860. https://doi.org/10.1002/rob.2154

Oey, M.L., Vanstreels, E., De Baerdemaeker, J., Tijskens, E., Ramon, H., Nicolaï, B., 2006. Influence of Turgor on Micromechanical and Structural Properties of Apple Tissue 1, 1879–1889. https://doi.org/10.1051/IUFoST:20060855

Olenskyj, A.G., Donis-González, I.R., Bornhorst, G.M., 2020. Nondestructive characterization of structural changes during in vitro gastric digestion of apples using 3D time-series micro-computed tomography. J. Food Eng. 267, 109692. https://doi.org/10.1016/j.jfoodeng.2019.109692

Olenskyj, A.G., Mennah-Govela, Y.A., Swackhamer, C., Rios-Villa, K.A., Bornhorst, G.M., 2017. Softening Half-Time and Final Normalized Hardness as Indicators of Food Structural Breakdown During In Vitro Digestion, in: Institute of Food Technologists. Las Vegas NV.

Opazo-Navarrete, M., Altenburg, M.D., Boom, R.M., Janssen, A.E.M., 2018. The Effect of Gel Microstructure on Simulated Gastric Digestion of Protein Gels. Food Biophys. 13, 124–138. https://doi.org/10.1007/s11483-018-9518-7

Panek, E., Gozdowski, D., 2021. Relationship between MODIS Derived NDVI and Yield of Cereals for Selected European Countries. Agronomy 11, 340. https://doi.org/10.3390/agronomy11020340

Parada, J., Aguilera, J.M., 2007. Food microstructure affects the bioavailability of several nutrients. J. Food Sci. 72, 21–32. https://doi.org/10.1111/j.1750-3841.2007.00274.x

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32.

Paula, A.M., Conti-Silva, A.C., 2014. Texture profile and correlation between sensory and instrumental analyses on extruded snacks. J. Food Eng. 121, 9–14. https://doi.org/10.1016/j.jfoodeng.2013.08.007

Peterlin, A., 1965. Diffusion in a network with discontinuous swelling. J. Polym. Sci. Part B Polym. Lett. 3, 1083–1087. https://doi.org/10.1002/pol.1965.110031222

Pinzer, B.R., Medebach, A., Limbach, H.J., Dubois, C., Stampanoni, M., Schneebeli, M., 2012. 3D-characterization of three-phase systems using X-ray tomography: Tracking the microstructural evolution in ice cream. Soft Matter 8, 4584–4594. https://doi.org/10.1039/c2sm00034b

Piper, D.W., Fenton, B.H., 1965. pH stability and activity curves of pepsin with special reference to their clinical importance. Gut 6, 506–8.

Pletschke, B.I., Naudé, R.J., Oelofsen, W., 1995. Ostrich pepsins I and II: A kinetic and thermodynamic investigation. Int. J. Biochem. Cell Biol. 27, 1293–1302. https://doi.org/10.1016/1357-2725(95)00092-4

Priori, S., Martini, E., Andrenelli, M.C., Magini, S., Agnelli, A.E., Bucelli, P., Biagi, M., Pellegrini, S., 2013. Improving Wine Quality through Harvest Zoning and Combined Use of Remote and Soil Proximal Sensing. https://doi.org/10.2136/sssaj2012.0376

Qiao, J., Wang, N., Ngadi, M.O., Kazemi, S., 2007. Predicting mechanical properties of fried chicken nuggets using image processing and neural network techniques. J. Food Eng. 79, 1065–1070. https://doi.org/10.1016/j.jfoodeng.2006.03.026

Qiu, D., Shao, S.X., Zhao, B., Wu, Y.C., Shi, L.F., Zhou, J.C., Chen, Z.R., 2012. Stability of ??-carotene in thermal oils. J. Food Biochem. 36, 198–206. https://doi.org/10.1111/j.1745-4514.2010.00526.x

Quevedo, R., Carlos, L.G., Aguilera, J.M., Cadoche, L., 2002. Description of food surfaces and microstructural changes using fractal image texture analysis. J. Food Eng. 53, 361–371.

https://doi.org/10.1016/S0260-8774(01)00177-7

Rahnemoonfar, M., Sheppard, C., 2017. Deep count: Fruit counting based on deep simulated learning. Sensors (Switzerland) 17, 1–12. https://doi.org/10.3390/s17040905

Raina, R., Madhavan, A., Ng, A.Y., 2009. Large-scale deep unsupervised learning using graphics processors. ACM Int. Conf. Proceeding Ser. 382. https://doi.org/10.1145/1553374.1553486

Ramlau, R., Scherzer, O., 2018. The first 100 years of the Radon transform. Inverse Probl. 34.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, 779–788. https://doi.org/10.1109/CVPR.2016.91

Redmon, J., Farhadi, A., 2018. YOLOv3: An Incremental Improvement.

Redmon, J., Farhadi, A., 2017. YOLO9000: Better, faster, stronger. Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 6517–6525. https://doi.org/10.1109/CVPR.2017.690

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Rizzolo, A., Vanoli, M., Cortellino, G., Spinelli, L., Herremans, E., Bongaers, E., Verboven, P., Nicolaï, B.M., Torricelli, A., 2013. Characterizing the tissue of apple air-dried rings as affected by fruit maturity assessed by TRS and osmosis treatment using X-ray CT measurement , and relationship with ring crispness. Insid. Symp. 24, 1–6.

Roberts, L.G., 1963. Machine Perception of Three-Dimensional Solids.

Rodgers, G., Kuo, W., Schulz, G., Scheel, M., Migga, A., Bikis, C., Tanner, C., Kurtcuoglu, V., Weitkamp, T., Müller, B., 2021. Virtual histology of an entire mouse brain from formalin fixation to paraffin embedding. Part 1: Data acquisition, anatomical feature segmentation, tracking global volume and density changes. J. Neurosci. Methods 364, 109354. https://doi.org/10.1016/j.jneumeth.2021.109354

Rodgers, G., Tanner, C., Schulz, G., Migga, A., Kuo, W., Bikis, C., Scheel, M., Kurtcuoglu, V., Weitkamp, T., Müller, B., 2022. Virtual histology of an entire mouse brain from formalin fixation to paraffin embedding. Part 2: Volumetric strain fields and local contrast changes. J. Neurosci. Methods 365, 109385. https://doi.org/10.1016/j.jneumeth.2021.109385

Rose, J.C., Kicherer, A., Wieland, M., Klingbeil, L., Töpfer, R., Kuhlmann, H., 2016. Towards automated large-scale 3D phenotyping of vineyards under field conditions. Sensors (Switzerland) 16, 1–25. https://doi.org/10.3390/s16122136

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern

Recognit. 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

Santos, T.T., de Souza, L.L., dos Santos, A.A., Avila, S., Santos, A.A. dos, Avila, S., dos Santos, A.A., Avila, S., Santos, A.A. dos, Avila, S., Souza, L.L. De, Avila, S., de Souza, L.L., dos Santos, A.A., Avila, S., 2020. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. Comput. Electron. Agric. 170, 1–22. https://doi.org/10.1016/j.compag.2020.105247

Schoeman, L., Williams, P., du Plessis, A., Manley, M., 2016. X-ray micro-computed tomography (µCT) for non-destructive characterisation of food microstructure. Trends Food Sci. Technol. 47, 10–24. https://doi.org/10.1016/j.tifs.2015.10.016

Searcy, S.W., Schueller, J.K., Bae, Y.H., Borgelt, S.C., Stout, B.A., 1989. Mapping of spatially variable yield during grain combining. Trans. Am. Soc. Agric. Eng. 32, 826–829. https://doi.org/10.13031/2013.31077

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2016. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. ICCV.

Shelat, K.J., Nicholson, T., Flanagan, B.M., Zhang, D., Williams, B.A., Gidley, M.J., 2014. Rheology and microstructure characterisation of small intestinal digesta from pigs fed a red meat-containing Western-style diet. Food Hydrocoll. 44, 300–308. https://doi.org/10.1016/j.foodhyd.2014.09.036

Shrestha, A., Mahmood, A., 2019. Review of deep learning algorithms and architectures. IEEE Access 7, 53040–53065. https://doi.org/10.1109/ACCESS.2019.2912200

Silver, D.L., Monga, T., 2019. In Vino Veritas: Estimating Vineyard Grape Yield from Images Using Deep Learning, in: Canadian AI. Springer International Publishing, pp. 212–224. https://doi.org/10.1007/978-3-030-18305-9_17

Simard, P.Y., Steinkraus, D., Platt, J.C., 2015. Best practices for convolutional neural networks applied to visual document analysis, in: Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. IEEE Comput. Soc, pp. 958–963. https://doi.org/10.1109/ICDAR.2003.1227801

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. 1–14.

Singh, H., Ye, A., Ferrua, M.J., 2015. Aspects of food structures in the digestive tract. Curr. Opin. Food Sci. 3, 85–93. https://doi.org/10.1016/j.cofs.2015.06.007

Somaratne, G., Ye, A., Nau, F., Ferrua, M.J., Dupont, D., Singh, R.P., Singh, J., 2020. Egg white gel structure determines biochemical digestion with consequences on softening and mechanical disintegration during in vitro gastric digestion. Food Res. Int. 138, 109782. https://doi.org/10.1016/j.foodres.2020.109782

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Stasenko, N., Savinov, M., Pukalchik, M., Somov, A., 2021. Deep Learning for Postharvest Decay Prediction in Apples, in: IECON 2021. IEEE. https://doi.org/10.1109/IECON48115.2021.9589498

Sudeep, K.S., Pal, K.K., 2017. Preprocessing for image classification by convolutional neural networks. 2016 IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. RTEICT 2016 - Proc. 1778–1781. https://doi.org/10.1109/RTEICT.2016.7808140

Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M., 2017. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 10553 LNCS, 240–248. https://doi.org/10.1007/978-3-319-67558-9_28

Sun, L., Gao, F., Anderson, M.C., Kustas, W.P., Alsina, M.M., Sanchez, L., Sams, B., McKee, L., Dulaney, W., White, W.A., Alfieri, J.G., Prueger, J.H., Melton, F., Post, K., 2017. Daily mapping of 30 m LAI and NDVI for grape yield prediction in California vineyards. Remote Sens. 9, 1–18. https://doi.org/10.3390/rs9040317

Swackhamer, C., Zhang, Z., Taha, A.Y., Bornhorst, G.M., 2019. Fatty acid bioaccessibility and structural breakdown from in vitro digestion of almond particles. Food Funct. 10, 5174–5187. https://doi.org/10.1039/c9fo00789j

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 07-12-June, 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Tabilo-Munizaga, G., Barbosa-Cánovas, G. V, 2004. Rheology for the food industry. J. Food Eng. 67, 147–156. https://doi.org/10.1016/j.jfoodeng.2004.05.062

Tan, M., Le, Q. V., 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. 36th Int. Conf. Mach. Learn. ICML 2019 2019-June, 10691–10700.

Tan, M., Le, Q. V, 2021. EfficientNetV2 : Smaller Models and Faster Training.

The National Agricultural Library, 2018. USDA National Nutrient Database for Standard Reference: Apples, raw, with skin.

Théroux-Rancourt, G., Jenkins, M.R., Brodersen, C.R., McElrone, A., Forrestel, E.J., Earles, J.M., 2020. Digitally deconstructing leaves in 3D using X-ray microcomputed tomography and machine learning. Appl. Plant Sci. 8, 1–9. https://doi.org/10.1002/aps3.11380

Trinh, L., Lowe, T., Campbell, G.M., Withers, P.J., Martin, P.J., 2013. Bread dough aeration dynamics during pressure step-change mixing: Studies by X-ray tomography, dough density and population balance modelling. Chem. Eng. Sci. 101, 470–477. https://doi.org/10.1016/j.ces.2013.06.053

Tukey, J.W., 1977. Exploratory Data Analysis, 2nd ed. Addison-Wesley Pub. Co., Reading, MA.

Turbin-Orger, A., Babin, P., Boller, E., Chaunier, L., Chiron, H., Della Valle, G., Dendievel, R., Réguerre, A.L., Salvo, L., 2015. Growth and setting of gas bubbles in a viscoelastic matrix imaged by X-ray microtomography: The evolution of cellular structures in fermenting wheat flour dough. Soft Matter 11, 3373–3384. https://doi.org/10.1039/c5sm00100e

Ullah, J., Takhar, P.S., Sablani, S.S., 2014. Effect of temperature fluctuations on ice-crystal growth in frozen potatoes during storage. LWT - Food Sci. Technol. 59, 1186–1190. https://doi.org/10.1016/j.lwt.2014.06.018

Valdez, P., 2020. Apple Defect Detection Using Deep Learning Based Object Detection For Better Post Harvest Handling 1–5.

Van Wey, A.S., Cookson, A.L., Roy, N.C., McNabb, W.C., Soboleva, T.K., Wieliczko, R.J., Shorten, P.R., 2014. A mathematical model of the effect of pH and food matrix composition on fluid transport into foods: An application in gastric digestion and cheese brining. Food Res. Int. 57, 34–43. https://doi.org/10.1016/j.foodres.2014.01.002

Vásquez, N., Magán, C., Oblitas, J., Chuquizuta, T., Avila-George, H., Castro, W., 2018. Comparison between artificial neural network and partial least squares regression models for hardness modeling during the ripening process of Swiss-type cheese using spectral profiles. J. Food Eng. 219, 8–15. https://doi.org/10.1016/j.jfoodeng.2017.09.008

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 2017-Decem, 5999–6009.

Verboven, P., Defraeye, T., Nicolai, B., 2018. Measurement and visualization of food microstructure: Fundamentals and recent advances, Food Microstructure and Its Relationship with Quality and Stability. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100764-8.00001-0

Verboven, P., Kerckhofs, G., Mebatsion, H.K., Ho, Q.T., Temst, K., Wevers, M., Cloetens, P., Nicolai, B.M., Quang, T.H., Temst, K., Wevers, M., Cloetens, P., Nicolaï, B.M., 2008. Three-Dimensional Gas Exchange Pathways in Pome Fruit Characterized by Synchrotron X-Ray Computed Tomography. Plant Physiol. 147, 518–527. https://doi.org/10.1104/pp.108.118935

Vicent, V., Verboven, P., Ndoye, F.T., Alvarez, G., Nicolaï, B., 2017. A new method developed to characterize the 3D microstructure of frozen apple using X-ray micro-CT. J. Food Eng. 212, 154–164. https://doi.org/10.1016/j.jfoodeng.2017.05.028

Vincent, J.F.V., 1989. Relationship between density and stiffness of apple flesh. J. Sci. Food Agric. 47, 443–462. https://doi.org/10.1002/jsfa.2740470406

Volz, R.K., Harker, F.R., Lang, S., 2003. Firmness Decline in `Gala' Apple during Fruit Development. J. Amer. Soc. Hort. Sci. 128, 797–802.

W. J. Chancellor, 1981. Substituting Information for Energy in Agriculture. Trans. ASAE 24, 0802–0807. https://doi.org/10.13031/2013.34341

Wang, D., Martynenko, A., 2016. Estimation of total, open-, and closed-pore porosity of apple slices during drying. Dry. Technol. 34, 892–899. https://doi.org/10.1080/07373937.2015.1084632

Wang, S., Chen, F., Wu, J., Wang, Z., Liao, X., Hu, X., 2007. Optimization of pectin extraction assisted by microwave from apple pomace using response surface methodology. J. Food Eng. 78, 693–700. https://doi.org/10.1016/j.jfoodeng.2005.11.008

Weng, W., Zhu, X., 2015. UNet: Convolutional Networks for Biomedical Image Segmentation, in: MICCAI.

Willemink, M.J., Noël, P.B., 2019. The evolution of image reconstruction for CT-from filtered back projection to artificial intelligence The first clinical CT scan took about 2185–2195.

https://doi.org/10.1007/s00330-018-5810-7

Wolpert, J.A., Vilas, E.P., 1992. Estimating Vineyard Yields: Introduction to a Simple, Two-Step Method. Am. J. Enol. Vitic. 43, 384–388.

Wooster, T.J., Day, L., Xu, M., Golding, M., Oiseth, S., Keogh, J., Clifton, P., 2014. Impact of different biopolymer networks on the digestion of gastric structured emulsions. Food Hydrocoll. 36, 102–114. https://doi.org/10.1016/j.foodhyd.2013.09.009

Wu, Y., Chen, Y., Yuan, L., Liu, Z., Wang, L., Li, H., Fu, Y., 2020. Rethinking Classification and Localization for Object Detection. CVPR 2020 10183–10192. https://doi.org/10.1109/CVPR42600.2020.01020

Wulfsohn, D., Zamora, F.A., Téllez, C.P., Lagos, I.Z., García-Fiñana, M., 2012. Multilevel systematic sampling to estimate total fruit number for yield forecasts. Precis. Agric. 13, 256–275. https://doi.org/10.1007/s11119-011-9245-2

Xu, F., Jin, X., Zhang, L., Chen, X.D., 2017. Investigation on water status and distribution in broccoli and the effects of drying on water status using NMR and MRI methods. Food Res. Int. 96, 191–197. https://doi.org/10.1016/j.foodres.2017.03.041

Yaeger, L., Lyon, R., Webb, B., 1997. Effective training of a neural network character classifier for word recognition. Adv. Neural Inf. Process. Syst. 807–813.

Yang, Q., Shi, L., Han, J., Zha, Y., Zhu, P., 2019. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. F. Crop. Res. 235, 142–153. https://doi.org/10.1016/j.fcr.2019.02.022

Yang, Z., Yabansu, Y.C., Al-Bahrani, R., Liao, W. keng, Choudhary, A.N., Kalidindi, S.R., Agrawal, A., 2018. Deep learning approaches for mining structure-property linkages in high contrast composites from simulation datasets. Comput. Mater. Sci. 151, 278–287. https://doi.org/10.1016/j.commatsci.2018.05.014

Ye, S., Li, B., Li, Q., Zhao, H.P., Feng, X.Q., 2019. Deep neural network method for predicting the mechanical properties of composites. Appl. Phys. Lett. 115. https://doi.org/10.1063/1.5124529

Zhao, Y., Takhar, P.S., 2017. Micro X-ray computed tomography and image analysis of frozen potatoes subjected to freeze-thaw cycles. LWT - Food Sci. Technol. 79, 278–286. https://doi.org/10.1016/j.lwt.2017.01.051

Zhou, X., Ampatzidis, Y., Lee, W.S., Agehara, S., 2021. Postharvest Strawberry Bruise Detection Using Deep Learning, in: ASABE Annial International Meeting. https://doi.org/10.13031/aim.202100458

Zoui, Zouid, I., R. Siret, E Mehninagic, Maury, C., M. Chevalier, F. Jourjon, 2010. Evolution of grape berries during ripening:investigations into the links between their mechanical properties and the extractability of their skin anthocyanins. J. Int. des Sci. la Vigne du Vin 44, 87–99.