Longitudinal CT scanning for explainable early detection of postharvest disorders: The 'Braeburn' browning case

Dirk Elias Schut[a,*], Rachael Maree Wood[b], Rob Schouten[c], Robert van Liere[a,d], Tristan van Leeuwen[a,e], and Kees Joost Batenburg[a,f]

[a]Computational Imaging Group, Centrum Wiskunde en Informatica (CWI), Science Park 123, 1098 XG Amsterdam, The Netherlands
[b]Horticulture and Product Physiology, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[c]Wageningen Food and Biobased Research, Bornse Weilanden 9, 6708 WG Wageningen, The Netherlands
[d]Visualization Group, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands
[e]Mathematisch Instituut, Utrecht University, Budapestlaan 6, 3584 CD Utrecht, The Netherlands
[f]Leiden Institute of Advanced Computer Science (LIACS), Leiden University, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
[*]Corresponding author. E-mail: dirk.schut@cwi.nl

Abstract

Longitudinal computed tomography (CT) datasets quantify the internal state of agricultural products in 3D over time, making them very suitable for studying the progression of postharvest disorders. In this paper, we present a workflow for developing early detection systems using longitudinal CT datasets. In our workflow, each CT-voxel is treated as a time series, which facilitates the analysis and detection of gradually progressing disorders in four different ways. Firstly, visualizing the difference between two CT scans provides higher contrast visualizations to study the progression of a disorder. Secondly, features derived from regional changes can be more discriminative than features derived at one time point over the whole product. Thirdly, detection setups can be simulated at different scanning times, making it possible to compare potential factory setups. Fourthly, longitudinally explainable artificial intelligence answers a specific research question about the decision-making of a neural network. The workflow is showcased on a dataset of CT scans of 80 ‘Braeburn’ apples that developed core browning. We provide a detailed visualization of the progression of core browning during and after controlled atmosphere (CA) storage. Moreover, a deep neural network classifier was developed that had more than 90% accuracy in the early detection of browning. The presented workflow provides powerful tools to study and automatically detect disorders that develop over time, thereby aiding in waste reduction and marketability.

Keywords: non-destructive testing (NDT), image registration, apple, internal browning, explainable artificial intelligence (XAI), integrated gradients

1 Introduction

Roughly one-third of the food produced for human consumption is wasted globally (Gustavsson et al. 2011). Sorting agricultural products based on non-destructive measurements can reduce waste and improve customer satisfaction (Nicolai et al. 2014). X-ray imaging can be used to determine the internal quality of a product, such as the presence of internal disorders (Kotwaliwale et al. 2014). The two main X-ray imaging methods are radiography, which is faster and cheaper, and computed tomography (CT), which has higher contrast and allows imaging in 3D. Radiographs can be simulated from CT scans (Hendriksen et al. 2021), so a CT dataset can be used to study both CT and radiography.

Many internal disorders develop gradually, so multiple scans over time may be acquired to study the progression of these disorders (Lammertyn et al. 2003; Jarolmasjed et al. 2016). A dataset is longitudinal when the same samples were measured multiple times. With longitudinal datasets, it is possible to separate the natural variance in the products from the changes over time (Tijskens et al. 2015). Image registration (Zitova and Flusser 2003) can be used to align the orientation of CT scans after acquisition. Image registration has been used to provide precise side-by-side views (Muziri et al. 2016) or to calculate how much a sample changed over time (Van Dael et al. 2019). Outside of agriculture, the difference between two images (a difference image) has been used for higher contrast visualizations of medical scans (Newatia et al. 2007) and for detecting supernovae in astronomy (Zackay et al. 2016).

For gradually developing disorders early detection is valuable. Early detection can limit infectious disorders from spreading such as fruit flies (Yang et al. 2006). Moreover, some foods are still consumable if a disorder is detected at an early stage, but not at a later stage, such as apples with bitterpit (Jarolmasjed et al. 2016), or pears with mealiness (Muziri et al. 2016). Core browning is a common (Argenta et al. 2021) internal disorder in apples that is induced by controlled atmosphere (CA) storage, and continues developing out-of-storage (Watkins and Mattheis 2019). Apples are typically sorted immediately after CA storage, and the browning may develop further in the time between sorting and consumption. Therefore, browning must be detected at a very early stage to avoid brown apples reaching the consumer.

In recent years, many authors have reported successes in the application of deep learning methods to detect postharvest disorders (Tempelaere et al. 2023a). Deep learning has been applied on CT scans of many crops such as apples (Tempelaere et al. 2024), pears (Van de Looverbosch et al. 2021; Tempelaere et al. 2023b), avocado (Andriiashen et al. 2023; Matsui et al. 2023), mango (Ansah et al.

2023) and passion fruit (Lu et al. 2023). Deep learning models are 'black box' models, meaning that their inner workings are too complex to interpret directly by humans. Explainable artificial intelligence (XAI) methods have been developed to gain insight into how neural networks interpret images (Buhrmester et al. 2021). Different XAI methods exist that highlight different parts of the image and it is challenging to evaluate how well they perform (Hooker et al. 2019). To counter the evaluation problem Sundararajan et al. (2017) formalized axioms that an XAI technique should follow, and derived a technique called integrated gradients (IG) directly from these axioms.

In this paper, we present a workflow for using longitudinal CT datasets to develop early detection systems for postharvest disorders. The key contribution of this paper is showing that image registration of the CT scans makes it possible to treat individual voxels as longitudinal data, and that this benefits the development of X-ray-based early detection systems in four ways. 1. Improved contrast of visualizations. 2. More discriminative features for quantitative analysis. 3. Early detection classifiers can be compared. 4. More interpretable explainable AI.

As an example case of our workflow, we studied core browning in 'Braeburn' apples. To capture the entire development of the disorder, 359 CT scans of apples were acquired before, during, and after CA storage. The severity of browning was scored at the end of the shelf life period using destructive visual inspection. With our workflow, we developed deep neural networks that were capable of scoring the browning severity of apples on a 1-4 scale, and performing early detection of browning on the day that apples were taken out of CA storage.

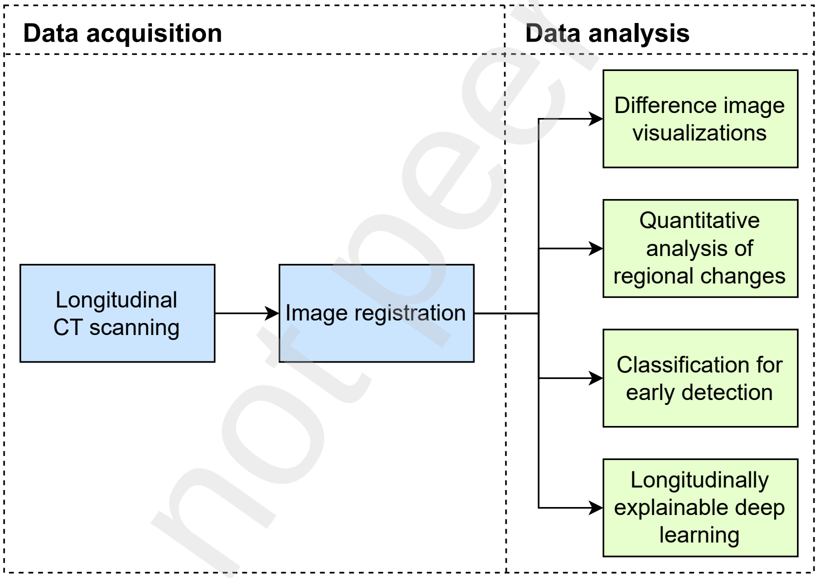

Figure 1: Overview of our workflow.

2 Workflow overview

Our workflow for developing early detection systems using longitudinal CT scanning consists of two data acquisition steps, and four methods that can be used to analyse the data (Figure 1). The ordering of the data analysis methods in Figure 1 relates to the different stages in developing an early detection system: First, visual inspection of the data is performed, then quantitative evaluation is performed, after that, an automatic detection system is developed, and finally, the automatic detection system is evaluated. We show in our results that image registration and longitudinal CT scanning provide added value in all these stages of development. A short description of each part of the workflow is provided below.

Longitudinal CT scanning. To acquire a longitudinal dataset, the same products must be scanned multiple times. The longitudinal CT dataset describes the progression of the disorder over time in 3D.

Image registration. Image registration should be used to align the orientation of the CT scans. After image registration, individual voxels can be treated as longitudinal data, which is required by the data analysis methods in our workflow.

Difference Image visualizations. A difference image results from subtracting two registered images of the same subject. The difference image will be zero on areas that are the same between both images and non-zero for areas that have changed. It can greatly improve the visibility of small changes on a static background (Newatia et al. 2007; Zackay et al. 2016). By subtracting the first CT scan from later CT scans the changes to the product can be visualized at each stage of the disorder.

Quantitative analysis of regional changes. Many features can be derived from CT datasets. The value of a feature varies because of the natural variation of an object, and because of the developments over time. By calculating the difference of a feature between two time points the effect of the natural variation can be eliminated (Tijskens et al. 2015). Moreover, if it is known that a disorder mostly affects a specific region features can be calculated only over that region. These two approaches can be combined to derive more discriminative features for quantitative analysis of disorders.

Classification for early detection. CT scans can be used to simulate arbitrary X-ray imaging geometries (Hendriksen et al. 2021). Classifiers (e.g. deep neural networks) can be trained on the simulated data. This makes it possible to compare potential factory X-ray imaging setups without having to acquire the same dataset multiple times with different hardware. With longitudinal datasets, it is possible to train classifiers to predict the state of the disorder at a later time point in the dataset.

Longitudinally explainable deep learning. Explainable AI (XAI) techniques aim to describe an AI's decision-making by scoring how much each input feature attributed to the AI's decision. The integrated gradients XAI technique (Sundararajan et al. 2017) can be used with a longitudinal baseline (Mamalakis et al. 2023) to answer the question: "How much did each change in the product over time affect the classification score over time?"

3 Methods

In this section, we describe how our workflow was applied in practice to a dataset of 'Braeburn' apples to develop an early detection system for core browning. All computational methods were implemented using scripts in the Python programming language. The Python code is publicly available on Github (`https://github.com/D1rk123/longitudinal_ct_workflow`).

3.1 Longitudinal CT scanning

3.1.1 Apple samples

In 2022, 80 'Braeburn' apple fruit were harvested at the optimal harvest date from the Randwijk experimental orchard, the Netherlands. To cause browning to develop, the apples were stored in controlled atmosphere (CA) storage with non-optimal conditions (0.5 °C, 1.5 kPa O_2, 5 kPa CO_2). The apples were stored in CA storage for 9, 12, or 17 weeks. After CA storage, the apples were stored in shelf-life conditions (19-20 °C) for up to three weeks. Varying durations of CA and shelf-life storage were used to ensure variety in the browning severity. Moreover, due to the scanning time and the availability of the scanner, not all apples were scanned at each interval. An overview of the scans is provided in Table 1. After the last CT scan the fruit were sliced with a knife into approximately 10 mm slices, photographed, and scored for browning intensity. Browning intensity was scored for each browning type on a subjective scale of 1 to 4, where; (1) indicates no brown tissue; (2) mild; (3) moderate and (4) severe browning intensity. One apple was excluded because it developed severe rot. A part of this dataset was used in a previous publication within the same research project (Wood et al. 2024).

Table 1: Overview of when the apples were scanned and sliced. When apples were sliced, they were scored on their level of browning.

CA storage (weeks)	**Shelf-life (days)**	**Scanned**	**Sliced**	**Healthy**	**Brown**	**Excluded (rot)**
0	0	32	-	-	-	-
3	0	10	1	1	0	0
9	0	21	-	-	-	-
	14	8	5	4	1	0
	21	3	3	1	2	0
12	0	30	3	3	0	0
	2	27	-	-	-	-
	9	27	-	-	-	-
	14	27	27	15	12	0
17	0	41	-	-	-	-
	2	41	-	-	-	-
	4	31	-	-	-	-
	7	41	21	5	15	1
	14	20	20	3	17	0
Totals		**359**	**80**	**32**	**47**	**1**

3.1.2 CT scanning protocol

The apples were scanned at the FleX-ray laboratory (Coban et al. 2020) using a custom scanner developed by TESCAN-XRE, Gent, Belgium. A cone beam geometry with a circular trajectory was used to acquire 1440 projection images at an exposure time of 100 ms, a tube peak voltage of 90 kV, and a current of 550 µA. A detector pixel binning of two was used and the voxel size was between 130.6 µm and 134.3 µm. The voxel sizes differed slightly between scans due to one scanner motor being defective on some scan days, which limited the range of possible resolutions but otherwise did not affect the scans. The average time per scan was five minutes.

3.1.3 Image reconstruction

The FDK algorithm (Feldkamp et al. 1984) was used for reconstructing the CT volumes. Third-order polynomial beam hardening correction was used (Herman 1979), using a paper cup filled with apple juice as a calibration phantom. Gradient descent with momentum was used to optimize the beam hardening parameters (Schoonhoven et al. 2024). The cost function was the mean squared error (MSE) between each voxel inside the juice and the mean value of that region plus the MSE between each voxel outside the cup and the mean value of that region. The gradient could be calculated automatically by using PyTorch autograd (Paszke et al. 2017) and the differentiable FDK implementation from the Tomosipo library (Hendriksen et al. 2021). The CT scan of a cup filled with juice was also used to normalize the CT gray values similarly to the porosity mapping method (Nugraha et al. 2019). In the remainder of the paper, a CT gray value of 1 represents a density equal to juice and a gray value of 0 represents a density equal to air.

3.2 Image registration

All CT scans were aligned in 3D to the first available CT scan of the same apple by using image registration with the SimpleITK library (Lowekamp et al. 2013; Avants et al. 2014). All apples were scanned with the stem on top, but the rotation around the stem-calyx axis differed between scans. To avoid converging to the wrong local minimum, the first optimization step of the registration method was performed ten times with a different initial rotation. The initial rotations were equally spaced between 0 and 360 degrees around the vertical axis. In this first step, a rigid transformation model was optimized over the MSE using the L-BFGS optimizer (D. C. Liu and Nocedal 1989) with four and two times downsampling. The image registration parameters with the lowest MSE were used as initialization for the second step. In the second step, a similarity transformation model was optimized over the MSE using the regular step gradient descent optimizer with four times, two times, and no downsampling.

3.3 Difference image visualizations

Difference images were calculated between the CT slices of the before-storage scan and all other scans of the same apple, to visualize which parts of the apple changed during CA storage. Difference images were also calculated from simulated radiographs. The radiographs had to be simulated because the image content of radiographs acquired at different angles is different, and therefore they could not be aligned with image registration. The radiographs were simulated from the registered CT scans using the Tomosipo Python library (Hendriksen et al. 2021).

3.4 Quantitative analysis of regional changes

3.4.1 Image segmentation

Image segmentation was performed to segment the apple into the core, inner apple, outer apple, and background. The aim was to be able to select the inner region as this is where core browning occurs and to separate it from the core and the outer regions. Figure 2 shows an example of the segmentation. The first available scan of each apple was segmented and these segmentation masks could also be used on the other scans of the same apple because all scans were aligned after the image registration.

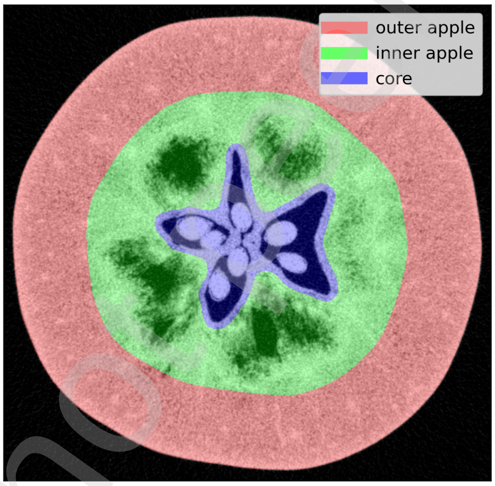

Figure 2: Example of the segmented regions.

To segment the full apple from the background, the CT scan was thresholded at 0.4. Morphological closing and hole filling were used as postprocessing to ensure the whole volume was included. To segment the apple core, only a central part of the CT scan was considered: 20 % above and below the center of mass vertically and the central 50 % of the image horizontally. Within this region, voxels below a threshold of 0.15 were selected as an initial estimate of the core. This mask was refined by using binary dilation, selecting the largest connected component, binary closing, and hole filling using Scikit-image (Walt et al. 2014). The outer apple region was defined as the region inside the apple at most 12 mm from the peel. The inner apple region was defined as the rest of the inside of the apple except the core region.

3.4.2 Porosity over time

The porosity is defined as the percentage of volume consisting of voids in a given region (Mendoza et al. 2007). CT scans can be interpreted as maps of local porosity using porosity mapping (Nugraha et al. 2019). To perform porosity mapping the CT scan needs to be corrected for beam hardening, and the range of values needs to be normalized to the density of juice, which was performed as described in Section 3.1.3. A local estimate of the porosity p of a voxel can be derived from the juice equivalent density d as follows:

$$p = 1 - d$$

The local porosity estimates were averaged over the whole apple and the inner apple region to estimate the porosity of these regions.

3.4.3 Statistical differences between browning scores

Statistical testing was performed to check if the porosity was significantly different for apples with different browning scores. Four porosity-related features were calculated: The whole apple and inner apple porosity on the last CT scan before slicing the apple, and their change from the before-storage scan. The features were calculated on the 25 apples for which the before-storage scan was available and which had been in CA storage for at least 12 weeks. Firstly, a one-way ANOVA test was performed with the null hypothesis that the means were the same for all browning scores. Secondly, a Tukey's HSD test was performed to test if the difference between individual browning scores was significant.

3.5 Classification for early detection

3.5.1 Comparing setups for detecting or predicting browning

Two tasks were investigated: to detect the browning score after shelf-life (detection), and to predict the future browning score after CA storage but before shelf-life (prediction). For both tasks, neural networks were trained on CT slices or real-time obtainable (100 ms) radiographs, so four neural networks were compared. Only apples that had been in CA storage for at least nine weeks were included. For the prediction networks, only apples that were sliced after two weeks of shelf life were included. A 60%, 20%, 20% split was used between the training, validation, and test sets. The exact number of images used can be found in Appendix A. Stratified sampling ensured each set contained a similar ratio of apples from each day and with each browning label. For the test set, if possible, apples were selected for which a before-storage scan was available, which was required for the explainable AI method described in Section 3.6.2.

3.5.2 Training

The neural network training was implemented using Pytorch (Paszke et al. 2017) and Pytorch Lightning (Falcon and team 2024). The network architecture was a slightly modified version of the efficientnetV2_s (Tan and Le 2021) implementation from torchvision (Marcel and Rodriguez 2010) pre-trained on ImageNet-1K (Russakovsky et al. 2015). The weights of the original three input channels (RGB-color) were averaged to obtain one grayscale input channel. Moreover, the network was converted to a regression network by replacing the linear layer and the softmax activation function at the output side of the network by a linear layer with one output. The cost function was the mean absolute error (MAE). The optimizer was ADAMW (Loshchilov and Hutter 2017). Data augmentations were applied using the Albumentations library (Buslaev et al. 2020) and the images were normalized based on the mean and standard deviation of the training set. The training duration was approximately 20000 epochs (2 days on 2 RTX Titan GPUs). At the end of every epoch, the MAE was calculated over the validation set, and the network weights were saved if the validation MAE was the lowest up to that point. The version of the network with the lowest validation MAE at the end of the training was used in the results section.

3.5.3 CT slices

For the networks trained on the CT scans, in every epoch, one horizontal slice was randomly selected from the region 10% above or below the center of mass of the core of the apple, resulting in on average 123 slices being available per apple. The data augmentation consisted of random rotations, flipping, shearing, translations, scaling, and elastic deformations (Simard et al. 2003).

3.5.4 Radiographs

The real radiographs that were acquired in the CT scanner were used for training the radiography neural networks. Flatfielding and the log-transformation were applied as preprocessing. To reduce storage space, only every third preprocessed radiograph was stored, resulting in 480 radiographs per apple. In every epoch, one radiograph was selected randomly per apple. The data augmentation consisted of random rotations, horizontal flipping, shearing, scaling, and elastic deformations (Simard et al. 2003). Moreover, 20% of the top and bottom of the radiograph were cropped to remove parts where browning does not occur.

3.5.5 Combining neural network outputs

For each apple multiple CT slices were available. These CT slices all had the same corresponding browning score. However, the scores outputted by the neural network were not always the same. The neural network outputs of all CT slices from one apple can be combined into a single score by taking the median. This way, errors in either direction cancel each other out, possibly improving the prediction of the browning score of the apple. This strategy was also applied to the radiographs. The MAE with the ground truth was calculated for all separate images and for the combined scores per apple. In both approaches, the score was rounded to the nearest integer before calculating the error.

3.6 Longitudinally explainable deep learning

3.6.1 Integrated gradients (IG)

To illustrate how the neural networks interpreted the X-ray images the integrated gradients (IG) (Sundararajan et al. 2017) explainable AI technique was used. The result of IG is a heatmap that highlights how much each part of the image has attributed to the browning score. To calculate this heatmap a second image is required that serves as a baseline. The choice of baseline image highly affects the outcome of IG (Sturmfels et al. 2020), and determines which research question is answered by the IG heatmap (Mamalakis et al. 2023).

There are several other techniques to calculate explainable AI heatmaps, but IG is unique in that it was designed to satisfy certain mathematical properties that make the heatmap more interpretable. Specifically, IG was designed to have the properties of sensitivity and implementation invariance. Sensitivity means that for every input and baseline that differ in one feature but have different predictions then the differing feature should be given a non-zero attribution. Implementation invariance means that the attributions are always identical for two functionally equivalent networks. Additionally, IG has the property of completeness, which means that the attributions add up to the difference between the output of (neural network) F at the input x and the baseline x' (Sundararajan et al. 2017).

3.6.2 Longitudinal baseline

Three baselines were compared in this paper. The constant black baseline (zero before normalization) was suggested in the original IG paper (Sundararajan et al. 2017). The constant zero baseline (zero after normalization) is the default of the Captum explainable AI library (Kokhlikyan et al. 2020), which was used to calculate the heatmaps. The registered before-storage scan was also used as a baseline (Mamalakis et al. 2023), which we will call the longitudinal baseline. Because of the completeness and sensitivity properties an IG heatmap shows how much the difference between the input and the baseline image affected the difference in neural network output between the input and baseline. For the longitudinal baseline this corresponds to the research question “How much did the changes in density from before storage affect the browning score?”

3.6.3 Augmentation based smoothing

To reduce noise and to avoid that the heatmap would be orientation dependent an augmentation-based smoothing procedure was applied similar to the method used by Gildenblat (2021). Multiple random affine and flipping augmentations were applied to both the baseline and the input image. The range of transformations was the same as during the training of the neural networks. Heatmaps were calculated for the transformed baselines and inputs, and then transformed back to the original space by applying the inverse transformation. One hundred heatmaps were averaged to obtain a smoothed heatmap.

3.6.4 Evaluating the baselines

The output of the neural network on the baseline should stay close to the 1-4 range of scores, which we define as being within the [0.5, 4.5] range. Otherwise, the heatmap will mostly explain the baseline instead of the input image. This was checked for all three baselines.

Additional checks were performed to evaluate if IG with a longitudinal baseline could answer the research question "How much did the changes in density from before storage affect the browning score?" Browning will only increase over time. Therefore, on apples that did develop browning (ground truth score > 1) the neural network output should be higher on the input image than on the baseline. On the apples that did not develop browning (ground truth score 1), the networks should also score the baseline as healthy (within the [0.5, 1.5] range). These properties were tested for ten images per apple per network. The means and standard deviations over the network outputs were also calculated separately for the brown and healthy images.

4 Results

4.1 Difference image visualizations

Figure 3 shows CT scans, simulated radiographs, and their difference images visualizing brown core changes within one representative apple over time. Lower-density regions developed on the CT scan in the region around the core. The density in these regions decreased more after the apple was taken out of storage, than during storage. On the destructive visual inspection after the last CT scan this apple had severe browning (score 4). All apples with browning followed a similar pattern.

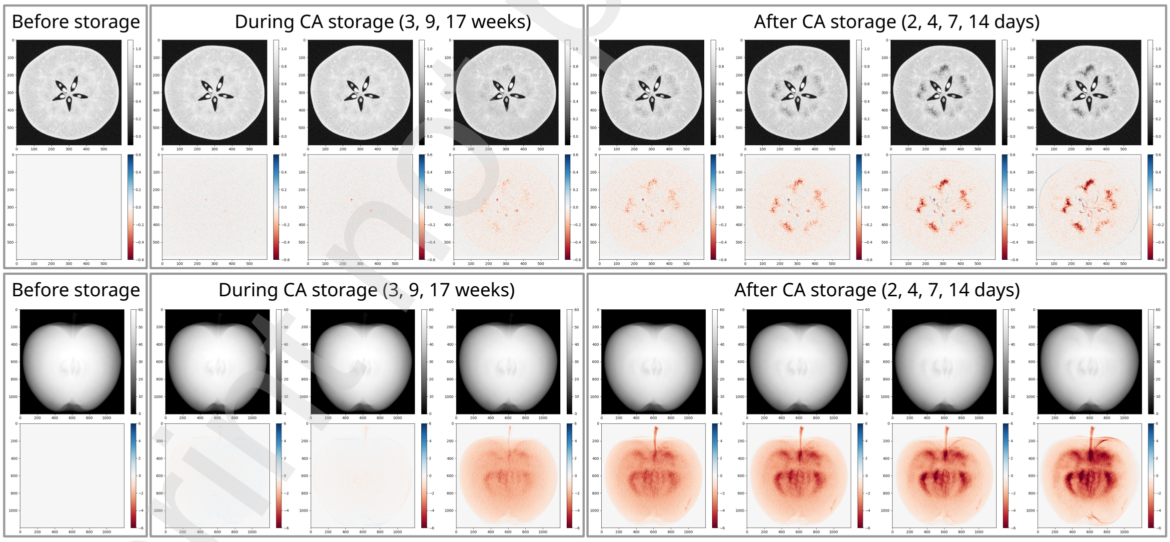

Figure 3: Top: Central slice of registered CT scans over time, and difference images showing the difference from the before storage scan. Bottom: Simulated radiographs of the same apple over time, and difference images showing the difference from the before storage radiograph.

4.2 Quantitative analysis of regional changes

Figure 4 shows how the whole apple porosity, change in whole apple porosity, inner apple porosity, and change in inner apple porosity relate to the browning score. The p-value of the one-way ANOVA test was highest on the whole apple porosity, and lowest on the change in inner apple porosity. The change in inner apple porosity also had the least overlap between the browning scores. It had a statistically significant ($p<0.05$) difference between scores 1, 2 and 3 (healthy, mildly brown and brown), and no significant difference scores 3 and 4 (brown and severely brown).

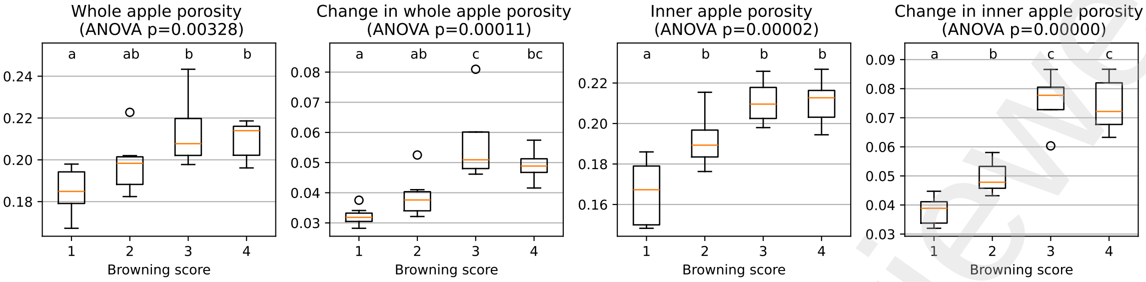

Figure 4: Boxplots of how four features related to porosity vary for each browning score. The orange line shows the median, the box shows the range between the first and the thrid quantile, and the whiskers show the range of values, except for outliers, which are shown as circles. Groups are statistically different ($p<0.05$) according to a Tukey's range test if they don't share any letters (abc). The p-value above each boxplot corresponds to a one-way ANOVA test with the null hypothesis that the means were the same for all browning scores.

4.3 Classification for early detection

The results of the neural networks on the test-set are shown in Table 2. To put these results into context, if a classifier had learned to output the same score for all apples, the score with both the lowest MAE and the highest yes/no accuracy would be 2. Only outputting this score would result in an MAE of 0.9333 on the detection task and 1.0 on the prediction task, and a yes/no accuracy of 0.667 on the detection task and 0.6 on the prediction task. All four neural networks outperform such a classifier, showing that they have learned to use the image data. The neural networks trained on CT slices performed better than the networks trained on radiographs in all cases. Combining the outputs of multiple input images also improved the results in most cases, resulting in a 100% detection and prediction of the presence of browning from CT scans. For the prediction task networks the validation set results were worse than the test-set results. The validation set MAE of the CT prediction network was 0.398 and 0.300 for separate and combined images respectively, and the yes/no accuracies were 81.2% and 90.0%. The validation set MAE of the radiograph prediction network was 0.825 and 0.700 for separate and combined images respectively, and the yes/no accuracies were 70.1% and 80.0%. The worst performing neural network was the radiograph prediction network. The training of this network also started overfitting earlier. The best MAE validation score was found at epoch 479, while for the other three networks, the best MAE validation score was found after more than 5000 epochs.

Table 2: Results of the neural networks on the test-set.

		Browning score MAE		**Browning yes/no accuracy**	
		Separate	**Combined**	**Separate**	**Combined**
CT slices	Detection	0.333	0.200	95.7%	100%
	Prediction	0.176	0.000	91.2%	100%
Radiographs	Detection	0.430	0.400	85.9%	86.7%
	Prediction	0.693	0.700	78.0%	80.0%

4.4 Longitudinally explainable deep learning

The neural network outputs on the zero and black baselines were far (14 or more) outside the 1-4 range of expected scores for all networks except the CT prediction network. On that network, the black baseline had a score of 0.94 and the zero baseline had a score of 1.14. Table 3 shows the results of the evaluation of the longitudinal baseline. The score was within the 1-4 range for all images on all networks and the other properties were met on most images. Figure 5 shows examples of IG with a longitudinal baseline for all four networks on the same apple.

Table 3: Results of the evaluation of the longitudinal baseline.

		Healthy mean and std. dev.	**Brown mean and std. dev.**	**Healthy and score 1**	**Brown and increasing**
CT slices	Detection	1.00 ± 0.03	1.12 ± 0.23	100%	100%
	Prediction	1.07 ± 0.07	1.55 ± 0.30	100%	90%
Radiographs	Detection	1.14 ± 0.20	1.54 ± 0.36	90%	98%
	Prediction	1.29 ± 0.05	1.29 ± 0.07	100%	77%

5 Discussion

5.1 Image registration

The image registration method in this paper is easy to reproduce because it is completely described in Section 3.2, has publicly available code, and relies on free software libraries only. Therefore it is a good starting point for image registration in future longitudinal X-ray imaging studies.

5.2 Difference image visualizations

Image registration makes it easier to compare images by showing them in the same orientation. When showing 2D slices of a 3D volume it also helps in showing the same part of the volume (Muziri et al. 2016; Schut et al. 2024). Figure 3 shows that difference images have the additional advantage of higher visual contrast, especially on the simulated radiograph images.

5.3 Quantitative analysis of regional changes

Figure 4 showed that using a region of interest, and looking at changes over time both resulted in more discriminative features, and that these two approaches can be combined to derive even more discriminative features. Calculating the change between two measurements is only one way of combining data in a longitudinal dataset. It is also possible to perform statistical analysis using measurements of more than two days by using non-linear mixed-effects regression (Tijskens et al. 2015). For example, non-linear mixed-effects regression was used to derive a biological age for ‘Granny Smith’ apples based on color measurements (Tijskens et al. 2008). Applying longitudinally derived features in sorting machines would require storing data for individual products through the supply chain, which is an upcoming trend in industry (Stephan et al. 2013).

The relation between porosity and browning was studied before by Chigwaya et al. (2021) in ‘Fiji’ apples. They found no statistically significant difference in whole apple porosity between apples that developed browning and apples that did not. However, when the porosity was calculated only over the coreline region the difference was significant. In earlier research within the same project, core browning was studied using the same dataset of ‘Braeburn’ apples, but the apples for which no before-storage scan was available were also included. For browning scores being at least two apart (e.g. 1 and 3) the density was significantly different, but for adjacent scores (e.g. 1 and 2) it was not (Wood et al. 2024). In Section (4.2) significant differences were found for the change in inner apple porosity between scores 1, 2, and 3, showing that the approach of data analysis on regional changes can result in more discriminative features.

5.4 Classification for early detection

The separate radiographs prediction setup is most similar to what currently could be implemented in a sorting system. Apples are typically sorted right after they come out of storage, and about ten radiographs could be acquired per second with an acquisition time of 100 ms, which is similar to the throughput in a commercial sorting system. The setups that combined multiple CT slices served as an indicator for the best-case performance if there is no limit on acquisition time. Moreover, research is ongoing to increase the throughput of CT scanners (Morton et al. 2009; De Schryver et al. 2016; Schut et al. 2022), so real-time CT scanning might become available in the future. The fact that a browning detection accuracy of 100% was reached on the combined CT slices setups suggests that

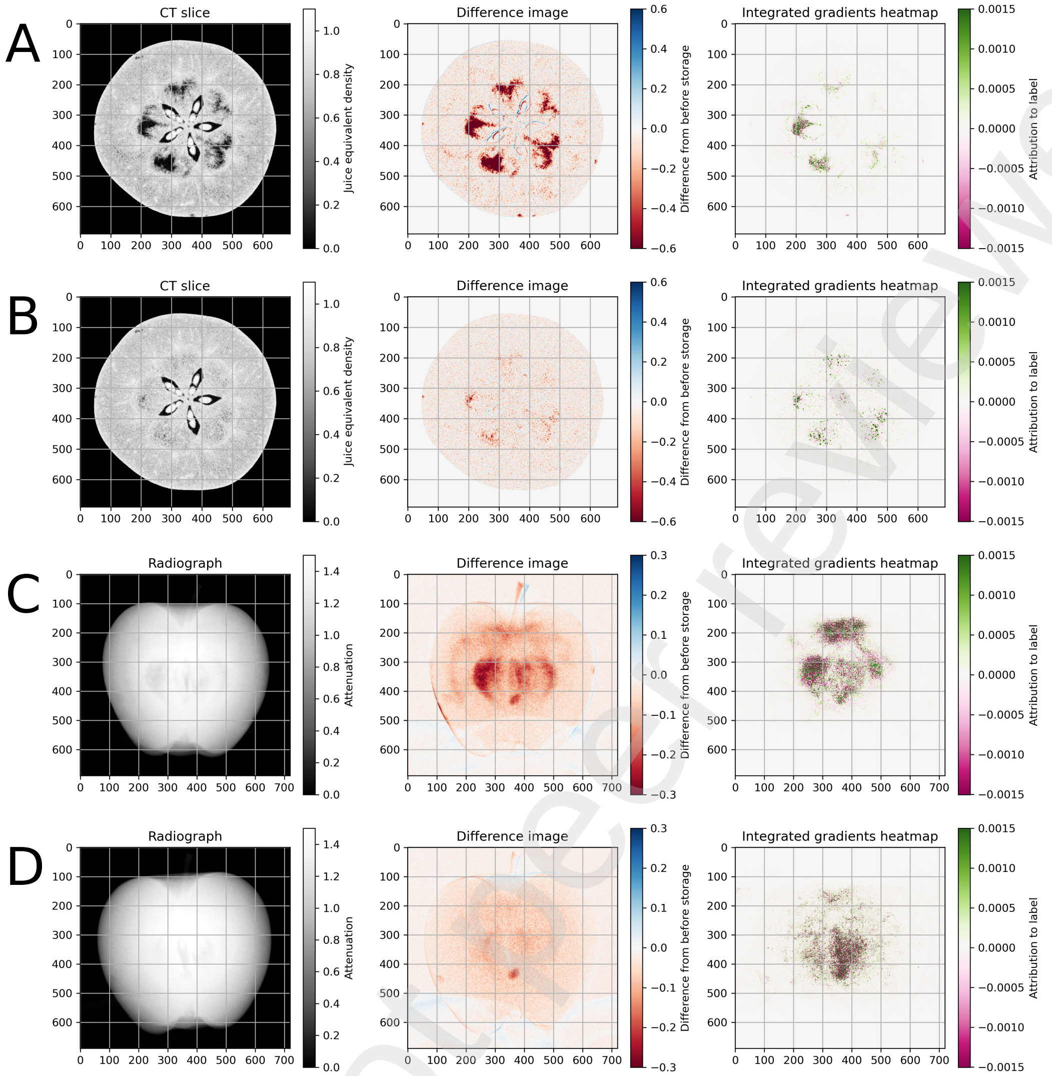

Figure 5: Examples of IG heatmaps with a longitudinal baseline A: Result on the CT detection network. B: Result on the CT prediction network. C: Result on the radiograph detection network. D: Result on the radiograph prediction network. All images were calculated on the same apple with browning score 4.

both detection and prediction of browning are possible if sufficiently high-quality X-ray images are available.

In our dataset, the grower, growing season, cultivar and storage conditions were the same for all apples. 'Braeburn' apples are typically very sensitive to browning disorders, but this also varies with orchard (Neuwald et al. 2014), and weather conditions (McCormick et al. 2021; Wood et al. 2022), which would need additional testing. On one hand, having more varied data could worsen the detection performance. On the other hand, more apples could be scanned to train the neural networks, which could improve the detection performance. The somewhat limited number of apples used for neural network training could also explain why the validation set results were better than the test-set results on the prediction task neural networks because the test and validation sets consisted of just 10 apples in the prediction task.

The result that browning can be detected on a single CT slice agrees with our earlier results (Wood et al. 2024). The IG heatmap is also similar to the void segmentation images that were used for browning detection in that paper. Tempelaere et al. (2024) also trained neural networks for detecting browning from CT scans and radiographs and achieved more than 95% accuracy with both. Our experiments extend upon these earlier works by scoring the severity of the browning on a 1-4 scale instead of yes/no classification, and by performing prediction of future browning from just-out-of-storage images. Another difference is that Tempelaere et al. (2024) used multiple CA storage conditions, with all apples from the browning-inducing CA storage conditions developing severe browning and all other apples remaining healthy. In our work, all apples were stored in the same CA conditions, and the resulting smaller changes between the brown and healthy apples are more challenging to classify and more in line with industry. J. He et al. (2024) compared dark-field and conventional transmission radiographs and suggested dark-field radiography is more suited to detect internal soft-tissue disorders like browning. Dark-field radiographs can not be simulated from transmission CT scans, but grating-based setups can capture transmission, dark-field, and phase-contrast CT scans at the same time (Bech et al. 2010) which could be used to acquire longitudinal multimodal CT datasets, adding another dimension for comparing setups.

5.5 Longitudinally explainable deep learning

Many baselines for IG have been suggested in the literature and the choice of baseline highly impacts the outcome of IG (Sturmfels et al. 2020; Mamalakis et al. 2023). Most research using IG has been done on classification networks, but in this paper, a regression network was used. Classification networks have a softmax or sigmoid activation function at the output, which are bounded to the [0, 1] range and have a high level of saturation (Rakitianskaia and Engelbrecht 2015), meaning that they map most of their inputs to values close to either zero or one. The regression networks used in this paper have a linear activation function at the output, which doesn't have this property. This could explain why the black and zero baselines did not meet the evaluation criteria on three of the four neural networks in this paper (Section 4.4). Furthermore, Sturmfels et al. (2020) showed that IG is blind to image features with the same color as the baseline. In the case of the X-ray images of apples, areas with browning show up as darker spots, with very brown spots in CT being almost black. Therefore a black baseline would be blind to these spots, which is undesirable. IG with a longitudinal baseline is blind to areas that did not change over time. There is background noise in all four difference images in Figure 5, but this does not show up in the heatmaps, showing that the neural networks are not activating on the background. Table 3 showed that most heatmaps did not contradict the fact that browning can only increase. This evaluation is not just an evaluation of the baseline, but also of the neural network (e.g. an untrained network with a perfect baseline image would not pass it).

The heatmaps in Figure 5 highlighted regions where browning was expected to occur. However, within these regions, there are both positive and negative values, often in neighboring voxels. It is unclear why the negative values appear within regions of interest, but it is a common phenomenon in IG heatmaps, so we don't expect it to be related to the apples. The standard approach to remove this phenomenon is to visualize the absolute value of the heatmap (Smilkov et al. 2017; Adebayo et al. 2018). However, with that approach the heatmaps lose the completeness property, making them less interpretable. Another approach, called XRAI (Kapishnikov et al. 2019), segments the image into small regions and postprocesses the IG heatmap so that each region has only one score. This removed the neighboring positive and negative values in the examples shown in that paper.

6 Conclusions

Longitudinal CT datasets quantify the internal state of agricultural products in 3D over time, making them very suitable for studying the progression of postharvest disorders. We presented a workflow for developing early detection systems using longitudinal CT data, and we showcased the workflow on a dataset of 'Braeburn' apples developing core browning. To align the CT scans of the same apple an image registration method was developed based on SimpleITK (Section 3.2), that registered the CT scans up to voxel precision (Figures 3 & 5). The development of core browning during and after CA storage on CT slices and radiographs was shown in detail (Figure 3). The method of difference images improved the visibility of the disorder, especially on the radiographs. The change in inner apple porosity was significantly different for apples with no browning, mild browning, and moderate browning (Figure 4). Calculating this feature only over the affected region (inner apple) and calculating the change between two time points made this feature more discriminative. Furthermore, our results suggest that by using neural networks on X-ray images browning can be scored on severity and future browning can be predicted from just-out-of-storage scans (Table 2). To find the optimal scanning setup to apply on a sorting line, many setups can be simulated from the same longitudinal CT dataset (Section 3.5). IG heatmaps were calculated to show how the neural network interpreted the images (Figure 5). Because we used a longitudinal baseline the heatmaps answer the research question “How much did the changes in density from before storage affect the browning score?” We conclude that by combining image registration and longitudinal CT scanning powerful data analysis techniques are made possible. Most of the presented methods use no apple-specific properties, so they should be applicable to other agricultural products or to longitudinal CT datasets from other fields.

Funding

This work was funded by the Dutch Research Council (NWO) through the UTOPIA project (ENWSS.2018.003).

Data availability

Data will be made available on request.

Appendices

A Number of images in the train, validation, and test sets

Table A.1: Division of apples into the train, validation, and test sets and the resulting number of CT slices and radiographs.

	Train set			Validation set			Test set		
	Apples	CT slices	Radiographs	Apples	CT slices	Radiographs	Apples	CT slices	Radiographs
Detection	45	5550	21600	15	1848	7200	15	1884	7200
Prediction	32	3922	15360	10	1240	4800	10	1248	4800

References

Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., and Kim, B. (2018). “Sanity checks for saliency maps”. In: *Advances in neural information processing systems* 31.

Andriiashen, V., Liere, R. van, Leeuwen, T. van, and Batenburg, K. J. (2023). “CT-based data generation for foreign object detection on a single X-ray projection”. In: *Scientific Reports* 13.1, p. 1881. DOI: https://doi.org/10.1038/s41598-023-29079-w.

Ansah, F. A., Amo-Boateng, M., Siabi, E. K., and Bordoh, P. K. (2023). “Location of seed spoilage in mango fruit using X-ray imaging and convolutional neural networks”. In: *Scientific African* 20. DOI: https://doi.org/10.1016/j.sciaf.2023.e01649.

Argenta, L. C., Freitas, S. T. de, Mattheis, J. P., Vieira, M. J., and Ogoshi, C. (2021). “Characterization and quantification of postharvest losses of apple fruit stored under commercial conditions”. In: *HortScience* 56.5, pp. 608–616. DOI: https://doi.org/10.21273/HORTSCI15771-21.

Avants, B. B., Tustison, N. J., Stauffer, M., Song, G., Wu, B., and Gee, J. C. (2014). “The Insight ToolKit image registration framework”. In: *Frontiers in neuroinformatics* 8, p. 44. DOI: https://doi.org/10.3389/fninf.2014.00044.

Bech, M., Jensen, T. H., Bunk, O., Donath, T., David, C., Weitkamp, T., Le Duc, G., Bravin, A., Cloetens, P., and Pfeiffer, F. (2010). “Advanced contrast modalities for X-ray radiology: Phase-contrast and dark-field imaging using a grating interferometer”. In: *Zeitschrift fuer medizinische Physik* 20.1, pp. 7–16. DOI: https://doi.org/10.1016/j.zemedi.2009.11.003.

Buhrmester, V., Münch, D., and Arens, M. (2021). “Analysis of explainers of black box deep neural networks for computer vision: A survey”. In: *Machine Learning and Knowledge Extraction* 3.4, pp. 966–989. DOI: https://doi.org/10.3390/make3040048.

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., and Kalinin, A. A. (2020). “Albumentations: Fast and Flexible Image Augmentations”. In: *Information* 11.2. DOI: https://doi.org/10.3390/info11020125.

Chigwaya, K., Karuppanapandian, T., Schoeman, L., Viljoen, D. W., Crouch, I. J., Nugraha, B., Verboven, P., Nicolai, B. M., and Crouch, E. M. (2021). “X-ray CT and porosity mapping to determine the effect of ‘Fuji’apple morphological and microstructural properties on the incidence of CO2 induced internal browning”. In: *Postharvest Biology and Technology* 174, p. 111464. DOI: https://doi.org/10.1016/j.postharvbio.2021.111464.

Coban, S. B., Lucka, F., Palenstijn, W. J., Van Loo, D., and Batenburg, K. J. (2020). “Explorative imaging and its implementation at the FleX-ray Laboratory”. In: *Journal of Imaging* 6.4, p. 18. DOI: https://doi.org/10.3390/jimaging6040018.

De Schryver, T., Dhaene, J., Dierick, M., Boone, M. N., Janssens, E., Sijbers, J., Dael, M. van, Verboven, P., Nicolai, B., and Van Hoorebeke, L. (2016). “In-line NDT with X-Ray CT combining sample rotation and translation”. In: *NDT & E International* 84, pp. 89–98. DOI: https://doi.org/10.1016/j.ndteint.2016.09.001.

Falcon, W. and team, T. P. L. (2024). *PyTorch Lightning*. DOI: 10.5281/zenodo.10779019. URL: https://doi.org/10.5281/zenodo.10779019.

Feldkamp, L. A., Davis, L. C., and Kress, J. W. (1984). “Practical cone-beam algorithm”. In: *Josa a* 1.6, pp. 612–619. DOI: https://doi.org/10.1364/JOSAA.1.000612.

Gildenblat, J. (2021). *PyTorch library for CAM methods*. https://github.com/jacobgil/pytorch-grad-cam.

Gustavsson, J., Cederberg, C., Sonesson, U., Van Otterdijk, R., and Meybeck, A. (2011). *Global food losses and food waste*. Food and Agriculture Organization of the United Nations (FAO).

He, J., Van Doorselaer, L., Tempelaere, A., Vignero, J., Saeys, W., Bosmans, H., Verboven, P., and Nicolai, B. (2024). “Nondestructive internal disorders detection of ‘Braeburn’apple fruit by X-ray dark-field imaging and machine learning”. In: *Postharvest Biology and Technology* 214, p. 112981. DOI: https://doi.org/10.1016/j.postharvbio.2024.112981.

Hendriksen, A., Schut, D., Palenstijn, W. J., Viganò, N., Kim, J., Pelt, D., Leeuwen, T. van, and Batenburg, K. J. (2021). “Tomosipo: Fast, Flexible, and Convenient 3D Tomography for Complex Scanning Geometries in Python”. In: *Optics Express*. DOI: https://doi.org/10.1364/oe.439909.

Herman, G. T. (1979). “Correction for beam hardening in computed tomography”. In: *Physics in Medicine & Biology* 24.1, p. 81. DOI: https://doi.org/10.1088/0031-9155/24/1/008.

Hooker, S., Erhan, D., Kindermans, P.-J., and Kim, B. (2019). “A benchmark for interpretability methods in deep neural networks”. In: *Advances in neural information processing systems* 32.

Jarolmasjed, S., Espinoza, C. Z., Sankaran, S., and Khot, L. R. (2016). “Postharvest bitter pit detection and progression evaluation in ‘Honeycrisp’apples using computed tomography images”. In: *Postharvest Biology and Technology* 118, pp. 35–42. DOI: https://doi.org/10.1016/j.postharvbio.2016.03.014.

Kapishnikov, A., Bolukbasi, T., Viégas, F., and Terry, M. (2019). “Xrai: Better attributions through regions”. In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4948–4957. DOI: https://doi.org/10.1109/iccv.2019.00505.

Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Alsallakh, B., Reynolds, J., Melnikov, A., Kliushkina, N., Araya, C., Yan, S., et al. (2020). “Captum: A unified and generic model interpretability library for pytorch”. In: *arXiv preprint arXiv:2009.07896*. DOI: https://doi.org/10.48550/arXiv.2009.07896.

Kotwaliwale, N., Singh, K., Kalne, A., Jha, S. N., Seth, N., and Kar, A. (2014). “X-ray imaging methods for internal quality evaluation of agricultural produce”. In: *Journal of food science and technology* 51, pp. 1–15. DOI: https://doi.org/10.1007/s13197-011-0485-y.

Lammertyn, J., Dresselaers, T., Van Hecke, P., Jancsók, P., Wevers, M., and Nicolaı, B. (2003). “Analysis of the time course of core breakdown in ‘Conference’pears by means of MRI and X-ray CT”. In: *Postharvest biology and technology* 29.1, pp. 19–28. DOI: https://doi.org/10.1016/s0925-5214(02)00212-0.

Liu, D. C. and Nocedal, J. (1989). “On the limited memory BFGS method for large scale optimization”. In: *Mathematical programming* 45.1, pp. 503–528. DOI: https://doi.org/10.1007/bf01589116.

Loshchilov, I. and Hutter, F. (2017). “Decoupled weight decay regularization”. In: *arXiv preprint arXiv:1711.05101*. DOI: https://doi.org/10.48550/arXiv.1711.05101.

Lowekamp, B. C., Chen, D. T., Ibáñez, L., and Blezek, D. (2013). “The design of SimpleITK”. In: *Frontiers in neuroinformatics* 7, p. 45. DOI: https://doi.org/10.3389/fninf.2013.00045.

Lu, Y., Wang, R., Hu, T., He, Q., Chen, Z. S., Wang, J., Liu, L., Fang, C., Luo, J., Fu, L., et al. (2023). “Nondestructive 3D phenotyping method of passion fruit based on X-ray micro-computed tomography and deep learning”. In: *Frontiers in Plant Science* 13, p. 1087904. DOI: https://doi.org/10.3389/fpls.2022.1087904.

Mamalakis, A., Barnes, E. A., and Ebert-Uphoff, I. (2023). “Carefully choose the baseline: Lessons learned from applying XAI attribution methods for regression tasks in geoscience”. In: *Artificial Intelligence for the Earth Systems* 2.1, e220058. DOI: https://doi.org/10.1175/aies-d-22-0058.1.

Marcel, S. and Rodriguez, Y. (2010). “Torchvision the machine-vision package of torch”. In: *Proceedings of the 18th ACM international conference on Multimedia*, pp. 1485–1488. DOI: https://doi.org/10.1145/1873951.1874254.

Matsui, T., Sugimori, H., Koseki, S., and Koyama, K. (2023). “Automated detection of internal fruit rot in Hass avocado via deep learning-based semantic segmentation of X-ray images”. In: *Postharvest Biology and Technology* 203, p. 112390. DOI: https://doi.org/10.1016/j.postharvbio.2023.112390.

McCormick, R., Biegert, K., and Streif, J. (2021). “Occurrence of physiological browning disorders in stored ‘Braeburn’apples as influenced by orchard and weather conditions”. In: *Postharvest Biology and Technology* 177, p. 111534. DOI: https://doi.org/10.1016/j.postharvbio.2021.111534.

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

Morton, E., Mann, K., Berman, A., Knaup, M., and Kachelrieß, M. (2009). “Ultrafast 3D reconstruction for X-ray real-time tomography (RTT)”. In: *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*. IEEE, pp. 4077–4080. DOI: https://doi.org/10.1109/NSSMIC.2009.5402393.

Muziri, T., Theron, K., Cantre, D., Wang, Z., Verboven, P., Nicolai, B., and Crouch, E. (2016). “Microstructure analysis and detection of mealiness in ‘Forelle’pear (Pyrus communis L.) by means of X-ray computed tomography”. In: *Postharvest Biology and Technology* 120, pp. 145–156. DOI: https://doi.org/10.1016/j.postharvbio.2016.06.006.

Neuwald, D. A., Sestari, I., Kittemann, D., Streif, J., Weber, A., and Brackmann, A. (2014). "Can mineral analysis be used as a tool to predict 'Braeburn' Browning Disorders (BBD) in apple in commercial controlled atmosphere (CA) storage in Central Europe". In: *Erwerbs-Obstbau* 56.1, pp. 35–41. DOI: https://doi.org/10.1007/s10341-014-0202-x.

Newatia, A., Khatri, G., Friedman, B., and Hines, J. (2007). "Subtraction imaging: applications for nonvascular abdominal MRI". In: *American Journal of Roentgenology* 188.4, pp. 1018–1025. DOI: https://doi.org/10.2214/ajr.05.2182.

Nicolai, B. M., Defraeye, T., De Ketelaere, B., Herremans, E., Hertog, M. L., Saeys, W., Torricelli, A., Vandendriessche, T., and Verboven, P. (2014). "Nondestructive measurement of fruit and vegetable quality". In: *Annual review of food science and technology* 5, pp. 285–312. DOI: https://doi.org/10.1146/annurev-food-030713-092410.

Nugraha, B., Verboven, P., Janssen, S., Wang, Z., and Nicolai, B. M. (2019). "Non-destructive porosity mapping of fruit and vegetables using X-ray CT". In: *Postharvest Biology and Technology* 150, pp. 80–88. DOI: https://doi.org/10.1016/j.postharvbio.2018.12.016.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. (2017). "Automatic differentiation in PyTorch". In: *NIPS-W*.

Rakitianskaia, A. and Engelbrecht, A. (2015). "Measuring saturation in neural networks". In: *2015 IEEE symposium series on computational intelligence*. IEEE, pp. 1423–1430. DOI: https://doi.org/10.1109/ssci.2015.202.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). "Imagenet large scale visual recognition challenge". In: *International journal of computer vision* 115, pp. 211–252. DOI: https://doi.org/10.1007/s11263-015-0816-y.

Schoonhoven, R., Skorikov, A., Palenstijn, W. J., Pelt, D. M., Hendriksen, A. A., and Batenburg, K. J. (2024). "How auto-differentiation can improve CT workflows: classical algorithms in a modern framework". In: *Optics Express* 32.6, pp. 9019–9041. DOI: https://doi.org/10.1364/oe.502920.

Schut, D. E., Batenburg, K. J., Liere, R. van, and Leeuwen, T. van (2022). "TOP-CT: Trajectory With Overlapping Projections X-Ray Computed Tomography". In: *IEEE Transactions on Computational Imaging* 8, pp. 598–608. DOI: https://doi.org/10.1109/TCI.2022.3192125.

Schut, D. E., Wood, R. M., Trull, A. K., Schouten, R., Liere, R. van, Leeuwen, T. van, and Batenburg, K. J. (2024). "Joint 2D to 3D image registration workflow for comparing multiple slice photographs and CT scans of apple fruit with internal disorders". In: *Postharvest Biology and Technology* 211, p. 112814. DOI: https://doi.org/10.1016/j.postharvbio.2024.112814.

Simard, P. Y., Steinkraus, D., Platt, J. C., et al. (2003). "Best practices for convolutional neural networks applied to visual document analysis". In: *Icdar*. Vol. 3. 2003. Edinburgh. DOI: https://doi.org/10.1109/icdar.2003.1227801.

Smilkov, D., Thorat, N., Kim, B., Viégas, F., and Wattenberg, M. (2017). "Smoothgrad: removing noise by adding noise". In: *arXiv preprint arXiv:1706.03825*. DOI: https://doi.org/10.48550/arXiv.1706.03825.

Stephan, P., Eich, M., Neidig, J., Rosjat, M., and Hengst, R. (2013). "Applying digital product memories in industrial production". In: *SemProM: Foundations of semantic product memories for the Internet of Things*. Springer, pp. 283–304. DOI: https://doi.org/10.1007/978-3-642-37377-0_17.

Sturmfels, P., Lundberg, S., and Lee, S.-I. (2020). "Visualizing the impact of feature attribution baselines". In: *Distill* 5.1. DOI: https://doi.org/10.23915/distill.00022.

Sundararajan, M., Taly, A., and Yan, Q. (2017). "Axiomatic attribution for deep networks". In: *International conference on machine learning*. PMLR, pp. 3319–3328.

Tan, M. and Le, Q. (2021). "Efficientnetv2: Smaller models and faster training". In: *International conference on machine learning*. PMLR, pp. 10096–10106.

Tempelaere, A., De Ketelaere, B., He, J., Kalfas, I., Pieters, M., Saeys, W., Van Belleghem, R., Van Doorselaer, L., Verboven, P., and Nicolai, B. M. (2023a). "An introduction to artificial intelligence in machine vision for postharvest detection of disorders in horticultural products". In: *Postharvest Biology and Technology* 206, p. 112576. DOI: https://doi.org/10.1016/j.postharvbio.2023.112576.

Tempelaere, A., Phan, H. M., Van De Looverbosch, T., Verboven, P., and Nicolai, B. (2023b). "Non-destructive internal disorder segmentation in pear fruit by X-ray radiography and AI". In: *Com-*

puters and Electronics in Agriculture* 212, p. 108142. DOI: https://doi.org/10.1016/j.compag.2023.108142.

Tempelaere, A., Van Doorselaer, L., He, J., Verboven, P., and Nicolai, B. M. (2024). "BraeNet: Internal disorder detection in 'Braeburn' apple using X-ray imaging data". In: *Food Control* 155, p. 110092. DOI: https://doi.org/10.1016/j.foodcont.2023.110092.

Tijskens, L., Konopacki, P., Schouten, R., Hribar, J., and Simčič, M. (2008). "Biological variance in the colour of Granny Smith apples: Modelling the effect of senescence and chilling injury". In: *Postharvest biology and technology* 50.2-3, pp. 153–163. DOI: https://doi.org/10.1016/j.postharvbio.2008.05.008.

Tijskens, L., Schouten, R., Konopacki, P., and Jongbloed, G. (2015). "Basic principles of analysing biological and technical variation in non-destructive data". In: *Computers and Electronics in Agriculture* 111, pp. 121–126. DOI: https://doi.org/10.1016/j.compag.2014.12.022.

Van Dael, M., Verboven, P., Zanella, A., Sijbers, J., and Nicolai, B. (2019). "Combination of shape and X-ray inspection for apple internal quality control: In silico analysis of the methodology based on X-ray computed tomography". In: *Postharvest Biology and Technology* 148, pp. 218–227. DOI: https://doi.org/10.1016/j.postharvbio.2018.05.020.

Van de Looverbosch, T., Raeymaekers, E., Verboven, P., Sijbers, J., and Nicolai, B. (2021). "Non-destructive internal disorder detection of Conference pears by semantic segmentation of X-ray CT scans using deep learning". In: *Expert Systems with Applications* 176, p. 114925. DOI: https://doi.org/10.1016/j.eswa.2021.114925.

Walt, S. van der, Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T., and contributors, the scikit-image (June 2014). "scikit-image: image processing in Python". In: *PeerJ* 2, e453. DOI: https://doi.org/10.7717/peerj.453.

Watkins, C. B. and Mattheis, J. P. (2019). "Apple". In: *Postharvest physiological disorders in fruits and vegetables*. Ed. by S. T. De Freitas, S. Pareek, S. De Freitas, and S. Pareek. CRC Press. Chap. 8, pp. 165–206. DOI: https://doi.org/10.1201/b22001-8.

Wood, R. M., Schut, D. E., Trull, A. K., Marcelis, L. F., and Schouten, R. E. (2024). "Detecting internal browning in apple tissue as determined by a single CT slice in intact fruit". In: *Postharvest Biology and Technology* 211, p. 112802. DOI: https://doi.org/10.1016/j.postharvbio.2024.112802.

Wood, R. M., Proske, M., Freitas, S. T. de, Scheer, C., Vögele, R. T., and Neuwald, D. A. (2022). "Seasonal variation in calcium and ascorbic acid content at harvest related to internal browning in 'Braeburn'apple during controlled atmosphere storage". In: *Scientia Horticulturae* 297, p. 110943. DOI: https://doi.org/10.1016/j.scienta.2022.110943.

Yang, E.-C., Yang, M.-M., Liao, L.-H., Wu, W.-Y., Chen, T.-W., Chen, T.-M., Lin, T.-T., and Jiang, J.-A. (2006). "Non-destructive quarantine technique-potential application of using X-ray images to detect early infestations caused by Oriental fruit fly (Bactrocera dorsalis)(Diptera: Tephritidae) in fruit". In: *Formosan Entomol* 26, pp. 171–186.

Zackay, B., Ofek, E. O., and Gal-Yam, A. (2016). "Proper image subtraction—optimal transient detection, photometry, and hypothesis testing". In: *The Astrophysical Journal* 830.1, p. 27. DOI: https://doi.org/10.3847/0004-637x/830/1/27.

Zitova, B. and Flusser, J. (2003). "Image registration methods: a survey". In: *Image and vision computing* 21.11, pp. 977–1000. DOI: https://doi.org/10.1016/S0262-8856(03)00137-9.