Journal of *Imaging*

MDPI

Article

Longitudinal CT Scanning for Explainable Early Detection of Postharvest Disorders: The 'Braeburn' Browning Case

Dirk Elias Schut [1,*], Rachael Maree Wood [2,3], Rob Schouten [4], Robert van Liere [1,5], Tristan van Leeuwen [1,6] and Kees Joost Batenburg [7]

1 Computational Imaging Group, Centrum Wiskunde en Informatica (CWI), 1098 XG Amsterdam, The Netherlands
2 Horticulture and Product Physiology, Wageningen University and Research, 6708 PB Wageningen, The Netherlands
3 Plant Production Systems Group, Wageningen University, 6708 PE Wageningen, The Netherlands
4 Wageningen Food and Biobased Research, 6708 WG Wageningen, The Netherlands
5 Visualization Cluster, Eindhoven University of Technology, 5612 AP Eindhoven, The Netherlands
6 Mathematisch Instituut, Utrecht University, 3584 CS Utrecht, The Netherlands
7 Leiden Institute of Advanced Computer Science (LIACS), Leiden University, 2333 CC Leiden, The Netherlands
* Correspondence: dirk.schut@cwi.nl

Abstract

This study presents two workflows for leveraging longitudinal computed tomography (CT) datasets when developing deep learning-based detection systems for gradually developing postharvest disorders. Workflow 1 (Longitudinal Benchmarking) benchmarks neural networks by training and testing them on images from different stages of disorder progression. It examines the trade-off between detecting a disorder early or accurately and evaluates whether neural networks can generalize across time points. Workflow 2 (Longitudinal eXplainable Artificial Intelligence (XAI) Heatmaps) provides heatmaps that indicate how changes over time affect the outcomes of neural networks. It uses image registration to align an earlier-acquired image and then uses it as a baseline when calculating the heatmap. The workflows are demonstrated on a dataset of 'Braeburn' apples that were CT-scanned multiple times while developing internal browning during controlled-atmosphere (CA) storage and shelf life. The Longitudinal Benchmarking workflow was used to investigate whether images acquired immediately after CA storage can be used to predict the eventual browning after a shelf-life period, which is highly relevant in industrial practice. Moreover, the longitudinal XAI heatmaps avoided artifacts caused by out-of-distribution baselines or identical baseline regions, which occurred with conventional black or zero baselines.

Keywords: non-destructive testing (NDT); deep learning; explainable artificial intelligence (XAI); image registration; Integrated Gradients; postharvest; Apple; internal browning

Check for updates

Academic Editor: Gabriel Cristobal

Received: 27 May 2026
Revised: 6 July 2026
Accepted: 13 July 2026
Published: 21 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Roughly one-third of the food produced for human consumption is wasted globally [1]. Sorting agricultural products based on non-destructive measurements can reduce waste and improve customer satisfaction [2]. Photography is a commonly used modality for postharvest disorder detection. However, it can only detect disorders that are visible on the product's exterior. To assess the internal quality of a product, X-ray imaging can be used [3]. The two main X-ray imaging methods are radiography, which is faster and cheaper, and computed tomography (CT), which provides higher contrast and enables 3D imaging.

Image classification algorithms can automatically assign a grade to each product on a sorting line based on the acquired images. Deep learning is a state-of-the-art technique for image classification, and it has been successfully applied for detecting disorders from CT or radiography images of many crops, such as apples [4], pears [5,6], avocado [7], and mango [8].

One challenge in detecting postharvest disorders is that many of them develop gradually. They may be difficult to detect at an early stage because they may not yet be clearly visible on the images provided to the neural network. However, detecting a disorder at an early stage may also be valuable, for example, to limit the spread of infectious disorders, such as fruit flies [9]. 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 bitter pit [10], or pears with mealiness [11]. To study disorder progression, multiple scans over time may be acquired of the same samples, forming a longitudinal dataset. Longitudinal CT datasets have been acquired to study several postharvest processes, such as internal browning in apples [12,13], bitter pit in apples [10], mealiness in pears [11], the preservation of strawberries and persimmon using edible coatings [14,15], and the aging of cucumbers [16]. However, to the knowledge of the authors, there is no prior work on deep learning-based classification of postharvest disorders that accounts for their progression over time.

Another challenge is that the decision-making of deep learning-based image classification networks is often unclear. The weights of the neural network after training fully specify the network's behavior, but they are impossible for humans to interpret directly due to the large number of weights and their complex structure. Explainable artificial intelligence (XAI) heatmaps (also called attribution maps or saliency maps) aim to provide a more easily understandable overview of a neural network's decision-making process. For a given neural network and input image, they show how much each region contributed to the network's decision [17–20]. However, there are many methods for calculating XAI heatmaps with strongly different results [21], and it can be unclear what a heatmap represents [22].

The key contribution of this paper is that we provide two workflows for using longitudinal imaging datasets in the development of deep learning-based image classification systems for postharvest disorders, which are both illustrated in Figure 1:

- Workflow 1: Longitudinal Benchmarking
- Workflow 2: Longitudinal XAI Heatmaps

The workflows only use longitudinal data during the development of classification systems. When the system is deployed, each product is scanned only once, and the image can be processed immediately, which is more practical for high-throughput industrial applications. Workflow 1 benchmarks neural networks by training and testing them on images from different stages of disorder progression. This makes it possible to investigate how much early-stage disorder detection affects accuracy and whether classification networks trained on data from one time point generalize to data from other time points. Workflow 2 provides heatmaps of how changes over time affect the outcome of neural networks. It avoids common artifacts of other XAI methods by only considering the differences between the input image and an earlier image of the same object.

The problem of detecting internal browning in apples of the 'Braeburn' cultivar is used as an example to demonstrate the workflows. We use a dataset from previous research [12,23], in which apples were CT scanned before, during, and after controlled-atmosphere (CA) storage, and afterwards scored for browning on a 1–4 scale. Radiographs can be simulated from CT scans [24], so we use this CT dataset to demonstrate the workflows on CT and radiography.

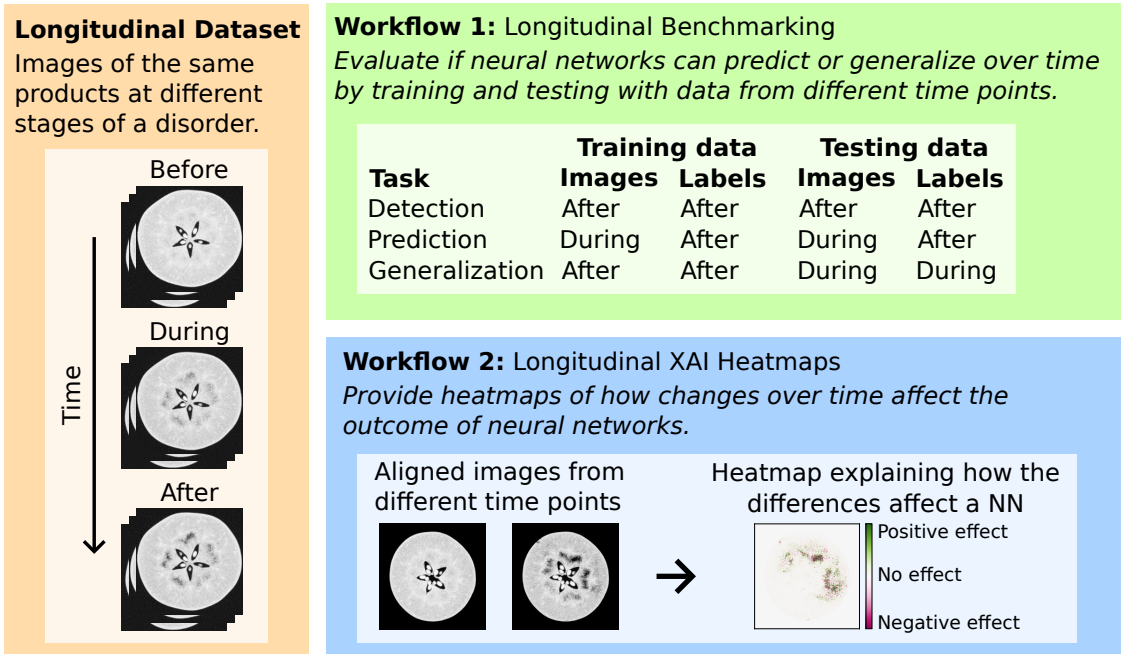


Figure 1. Illustration of the two workflows in this paper.

2. Related Work

2.1. Internal Browning in Apples

Internal browning (IB) is a disorder in apples and pears that is characterised by brown regions inside the fruit with an unpleasant taste. IB develops during controlled-atmosphere (CA) storage, which is used to store apples or pears for long periods to supply markets with a high-quality product year-round [25]. Browning occurs when cellular integrity breaks down, leading to the oxidation of phenolic compounds [26]. During this process, the affected tissues can also leak their intracellular fluids, which can then move towards the surrounding healthy tissues [27]. These changes in fluid distribution are visible on X-ray images due to the resulting differences in density. Regions with browning typically have a lower density, while surrounding tissues may have a higher density [27,28]. Depending on the apple cultivar, storage conditions, and several other pre- and post-harvest factors, different patterns of browning may occur [23,28]. Apples of the 'Braeburn' cultivar are especially susceptible to core browning [29,30], which affects the tissue around the core of the apple while the region closer to the peel remains unaffected, making non-destructive visual detection of this disorder difficult.

2.2. Postharvest Longitudinal CT Scanning

In the postharvest domain, longitudinal CT studies have been performed to study the progression of disorders [10–13,31–33], the spread of induced bacterial, fungal, or insect infections [9,34], aging [16,35], and the effects of coatings on long-term storage [14,15,35]. These datasets were investigated visually, or quantitative metrics were derived based on manual annotations or image processing techniques. Most studies scanned the whole fruit or vegetable, but some studies used high-resolution micro-CT scans of a small sample [36,37]. In several studies [11,13,31,36,37], scans of the same product were spatially aligned using image registration methods [38] to enable more precise comparisons between images or further processing.

On the topic of internal browning, a longitudinal CT dataset of pears was used to study internal browning and to model the gas transport inside the fruit [31]. A paper on detecting disorders by combining information from radiographs and 3D surface scans collected a

longitudinal CT dataset of apples as a reference dataset [13]. That dataset does not contain annotations or labels for specific disorders, but the paper shows an apple that develops core browning as an example. An earlier study from our research project investigated a longitudinal CT dataset of 'Braeburn' apples with browning labels to investigate how browning progresses before, during, and after CA storage [12]. That study found that the CT scans of apples that eventually would develop internal browning had a lower mean grey value at about 12 weeks of CA storage at a significance of $p < 0.1$. These differences became more significant ($p < 0.05$) when the CA storage was 18 weeks, or when a shelf-life period was added after CA storage. The experiments in this paper use the same dataset and extend the previous study by classifying individual apples using deep learning. To the authors' knowledge, this is the first study to apply deep learning to a longitudinal postharvest CT dataset.

2.3. AI-Based Postharvest Disorder Detection

A recently written introduction paper provides an overview of the research on using artificial intelligence for detecting postharvest disorders [39]. Image classification neural networks trained on radiography or CT datasets have been used to detect various types of postharvest disorders [4–8]. Deep learning-based image segmentation [40], object detection [41], and outlier detection [42] approaches have also been used.

To detect internal browning in 'Braeburn' apples, Tempelaere et al. [4] developed deep learning-based classifiers for CT scans and radiographs, and they achieved more than 95% accuracy on both types of data. However, they used different CA conditions for the apples that were labeled as brown or healthy, and they stored the apples in shelf-life conditions for 3 days before CT scanning, which together resulted in relatively large differences between the brown and healthy apples. In our dataset, the same storage conditions were used for all apples, and in the prediction task from Workflow 1 (Section 3.2), the apples are classified based on data from the day the apples were removed from CA storage. This is more in line with industrial practice, in which all apples are stored in the same CA conditions, sorted shortly after leaving CA storage, and consumed after a shelf-life period.

2.4. Explainable AI (XAI) Heatmaps

XAI heatmaps can be calculated in many ways, for example, by using gradients [17,19,43,44], internal activations of the network [20,45], local surrogate models [46], input perturbations [47–49], and propagating scores through the network back to the inputs [50,51]. Heatmaps calculated by different methods on the same image can vary a lot [21], and users commonly misinterpret heatmaps in practice [22].

To more precisely specify what heatmaps should represent, Sundararajan et al. [17] introduced the concept of axiomatic heatmap methods. Axiomatic heatmap methods are designed to satisfy appealing axioms (guaranteed mathematical properties), such as retaining linearity or symmetry of the neural network. Sundararajan et al. also introduced Integrated Gradients (IG) as the first axiomatic method [17], and the axioms of IG have been further investigated [18]. Several other XAI methods have been developed that satisfy many of the same axioms as IG [19,44,47,48].

One challenge in using IG and many other axiomatic XAI heatmap methods is that they require a baseline image as a second input. This baseline image is used as a neutral reference point in many of the axioms. The baseline image has a high impact on the result [52], and it may be challenging to find a suitable neutral baseline for a given problem [53,54]. In Workflow 2 (Section 3.3), we propose using an image from an earlier time point as a baseline, where we assume that in this image, the disorder is not yet present or less developed. This way, the baseline represents an object-specific neutral state instead of a general neutral

state. Mamalakis et al. [54] also used measurements from an earlier time point as a baseline to explain changes over time in a global climate model. Our Workflow 2 (Section 3.3) is similar, but it extends that work by using image registration to align the baseline with the input. Moreover, we investigate how this approach can be used to avoid artifacts related to out-of-distribution data [19], and regions that are the same in the baseline and input [52,53].

Previous research in postharvest disorder detection has used different XAI heatmap methods. The GradCAM method [20], used in [4,55,56], has a very low resolution because of how it is calculated. The LIME method [46], used in [57], only explains the behavior of the neural network in a region around the input. Moreover, it assigns the scores to superpixels (regions from a segmentation step), instead of to individual pixels, which reduces visual noise but limits the effective resolution. Workflow 2 provides full-resolution heatmaps with a well-defined mathematical foundation that explain how the neural network interpreted all changes between the baseline and input images.

3. Materials and Methods

3.1. Data Acquisition

Both workflows require a longitudinal dataset of images. How the longitudinal 'Braeburn' browning dataset [12,23] was collected and reconstructed is described below.

3.1.1. Apple Samples and Browning Score

In 2022, 80 'Braeburn' apple fruit were harvested at the optimal harvest date from the Randwijk experimental orchard, The Netherlands. To cause internal browning to develop, the apples were stored in controlled-atmosphere (CA) storage under non-optimal conditions (0.5 °C, 1.5 kPa O_2, 5 kPa CO_2). The apples were stored in CA storage for 0, 3, 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 limited scanner availability and long scanning times, not all apples were scanned at each interval. 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 based on visual inspection of the slices on a subjective scale from 1 to 4, where 1 indicates no brown tissue and 2, 3, and 4 indicate mild, moderate, and severe browning, respectively. One apple was excluded because it developed severe rot.

3.1.2. CT Scanning Protocol

The apples were scanned at the FleX-ray laboratory 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 [58] was used for reconstructing the CT volumes. Third-order polynomial beam hardening correction was used [59], using a paper cup filled with apple juice as a calibration phantom. Gradient descent with momentum was used to optimize the beam hardening parameters [60] by using the differentiable FDK implementation from the Tomosipo library [24].

3.2. Workflow 1: Longitudinal Benchmarking

3.2.1. Conceptual Overview of the Workflow

This workflow evaluates how much the performance of image classification networks is affected by the progression of the disorder. We distinguish three main tasks. For each task, neural networks are trained and tested on different combinations of input images and ground truth labels. **Detection**: Training and testing use images and labels from the same time point. This is the standard scenario for detection problems, and this does not require a longitudinal dataset. **Prediction**: Training and testing use images from one time point, but labels from a later time point, so the network should predict a future state of the product. **Generalization**: Networks are tested on data from a different time point than they were trained on.

3.2.2. Detection, Prediction and Generalization on the 'Braeburn' Browning Dataset

Four networks were trained to demonstrate the workflow on the 'Braeburn' browning dataset. The input data modality was either CT slices or radiographs. For both modalities, one network was trained to detect the browning score after shelf-life, and another was trained to predict the future browning score after CA storage but before shelf-life. Only apples that had been in CA storage for at least nine weeks were included. For the prediction task, only apples that were sliced after two weeks of shelf life were included. An overview of which apples were used for the detection and prediction tasks is provided in Table 1.

Table 1. Overview of which apples were used for the detection and prediction tasks.

		9 Weeks CA		**12 Weeks CA**	**17 Weeks CA**	
		14 Days Shelf-Life	**21 Days Shelf-Life**	**14 Days Shelf-Life**	**7 Days Shelf-Life**	**14 Days Shelf-Life**
Detection	Healthy	4	1	15	5	3
	Brown	1	2	12	15	17
Prediction	Healthy	3	N.A.	15	N.A.	3
	Brown	1	N.A.	12	N.A.	17

To evaluate the networks' ability to generalize to data from different time points, the detection networks were also tested on the data of the prediction task and vice versa.

3.2.3. Neural Network Architecture

The network architecture was an EfficientNetV2_s [61] using the implementation from torchvision [62] for all networks. This network architecture was chosen because it has shown a good performance relative to the training time and network size when compared to other networks [61]. The networks were converted for grayscale inputs by replacing the original three input channels (RGB-color) with one grayscale input channel. Moreover, the networks were converted to regression networks by replacing the linear layer and the softmax activation function at the output side of the networks with a single-output linear layer. By converting the networks into regression networks, the browning score was represented as a single continuous variable rather than four distinct classes. This more accurately reflects the ordered nature of the browning scores. Moreover, it benefits the calculation of XAI heatmaps in Workflow 2.

3.2.4. Usage of the Dataset

A 60%, 20%, 20% split was used between the training, validation, and test sets. The exact number of images used is given in Table 2. 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 Workflow 2 (Section 3.3). Data augmentations were applied using the Albumentations library [63] and the images were normalized based on the mean and standard deviation of the training set.

For the networks trained on the CT slices, in each 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, on average, in 123 slices being available per apple. The data augmentation consisted of random rotations, flipping, shearing, translations, scaling, and elastic deformations [64].

The radiography neural networks were trained on the radiographs that were also used as the raw measurement data for reconstructing the CT scans. Flatfielding and the log-transformation were applied as preprocessing steps. To reduce storage space, only every third preprocessed radiograph was stored, resulting in 480 radiographs per apple. In each epoch, one radiograph was randomly selected per apple. The data augmentation consisted of random rotations, horizontal flipping, shearing, scaling, and elastic deformations [64]. Moreover, roughly 20% of the top and bottom of the radiograph were cropped to remove parts where browning does not occur.

Table 2. 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	21,600	15	1848	7200	15	1884	7200
Prediction	31	3798	14,880	10	1240	4800	10	1248	4800

3.2.5. Neural Network Training

The neural network training was implemented using Pytorch (version 2.2.0) [65] and Pytorch Lightning [66]. The cost function was the mean absolute error (MAE). The optimizer was ADAMW [67], with learning rates of 1×10^{-5} and 2×10^{-5}, weight decays of 0.05 and 0.13, and batch sizes of 20 and 22 for the CT and radiography networks, respectively. No learning rate scheduling was used. The training duration was approximately 20,000 epochs (2 days on 2 RTX Titan GPUs). At the end of each epoch, the MAE was calculated on the validation set. The network weights with the lowest validation MAE were used in the Section 4.

3.2.6. Evaluation Metrics

To evaluate the neural networks, two metrics were calculated. For both metrics, the continuous neural network output was converted to a browning score by rounding to the nearest integer and then clipping to the 1–4 range. The first metric was the MAE of the browning score after rounding and clipping. The second metric is the percentage of apples correctly classified as brown or not, where browning scores 2, 3, and 4 are considered brown, and score 1 is considered healthy. This second metric was included for easier comparison with other papers that do not consider multiple levels of browning.

For each apple, multiple CT and radiography images were available, but the neural networks could return different scores for each image. The median score per apple was used to reduce errors and improve browning predictions. In the results, we report both per-image and per-apple scores.

3.3. Workflow 2: Longitudinal XAI Heatmaps

3.3.1. Conceptual Overview of the Workflow

This workflow provides heatmaps of how changes over time affect the outcome of neural networks. It works by applying any existing XAI heatmap method that satisfies the Completeness axiom while using an earlier image of the same product as the baseline, which we will call a longitudinal baseline. Image registration is used to spatially align the longitudinal baseline to the input. Using a longitudinal baseline is likely to avoid common issues, where the baseline is out-of-distribution or where the baseline is equal to the input image in important regions.

3.3.2. The Integrated Gradient (IG) Method

We demonstrated the workflow using the Integrated Gradients (IG) method [17]. For a given neural network F, input image $\bar{x}$ and baseline image x' the IG heatmap $A^{\mathrm{IG}}(\bar{x}, x', F)$ was calculated as follows:

$$A^{\mathrm{IG}}(\bar{x}, x', F) = (\bar{x} - x') \int_{\zeta=0}^{1} \nabla F((1-\zeta)x' + \zeta\bar{x}) d\zeta. \quad (1)$$

IG satisfies the Completeness axiom. This axiom guarantees that the sum over all pixels of the heatmap is equal to the difference in neural network output between the input and the baseline:

$$\sum_{i=1}^{n} A_i^{\mathrm{IG}}(\bar{x}, x', F) = F(\bar{x}) - F(x'). \quad (2)$$

By using an image that was acquired at an earlier time point as the baseline, the heatmap shows how much the changes over time affect the neural network's outcome.

3.3.3. Explaining the 'Braeburn' Browning Networks

To demonstrate the workflow on the 'Braeburn' browning dataset, we calculated heatmaps on how much each part of the image contributed to the browning score. We used the before-storage data as a longitudinal baseline, which should be unaffected by browning, as browning develops during storage. For the CT networks, the corresponding slice from the registered before-storage CT scan was used. For the radiography networks, a corresponding radiograph was simulated from the registered CT scan.

3.3.4. 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 [68,69]. 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 [70] 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.

3.3.5. Evaluating the Longitudinal Baseline

The workflow was evaluated on the 'Braeburn' browning dataset by comparing the longitudinal baseline with two other common baseline choices: the constant black baseline (zero before normalization), which was suggested in the original IG paper [17], and the

constant zero baseline (zero after normalization), which is the default of the Captum XAI library [71].

One property that a baseline should have is that it has a neutral output [17]. On the networks trained for browning detection, a neutral output is a browning score of one, so $F(x') = 1$. In that case, using the Completeness axiom (Equation (2)), the value of each pixel of the heatmap (A_i^{IG}) can be interpreted as how much that pixel contributed towards increasing the browning score from 1. When the baseline output has a large value, the heatmap will mostly explain the baseline and not the input. Therefore, the output values of the neural network on the considered baseline options were calculated, and they should be close to 1. Out-of-distribution data points may have high output values because the network was not optimized for these points during training.

Another property that a baseline should have is that it should not be equal to the input image in regions that correspond to the disorder. Due to how IG heatmaps are calculated, they are guaranteed to have a value of zero for pixels that are equal between the baseline and the input ($\bar{x}_i = x'_i$). This makes IG heatmaps blind to these regions [52,53]. Regions of severe browning may show up as black spots on a CT scan [28], so IG with a black baseline is blind to these regions, which is undesirable when trying to explain what contributes to a browning score. IG with a longitudinal baseline is blind to parts of the image that did not change over time. If you assume that apples are completely free of browning before storage, this is not a problem. The heatmaps were visually inspected for regions where they were equal to their inputs, and the heatmaps were checked for blind spots.

4. Results

4.1. Workflow 1: Longitudinal Benchmarking

The results of the neural networks on the test set are shown in Table 3. The neural networks trained on CT slices had better MAE and accuracy scores than the networks trained on radiographs in all cases. Combining the outputs from multiple input images of the same apple also improved results in all cases, resulting in 100% accuracy in the detection and prediction of browning from CT slices.

To put these results into context, consider a classifier that always predicts the same score. The score yielding the best results would be 2, with an MAE of 0.9333 on the detection task and 1.0 on the prediction task, and brown/healthy accuracies of 66.7% and 60%, respectively. All four neural networks strongly outperform such a classifier, indicating that they successfully learned to extract relevant information from the image data.

Table 4 shows the per-image MAE when the detection neural networks were applied to the prediction data and vice versa. It shows that the neural networks did not generalize across time points. Instead, they classified most apples at one side of the score range. The detection networks on the just-out-of-storage data classify most apples as healthy (score 1), while the prediction networks on the after-shelf-life data classify most apples as severely brown (score 4).

Table 3. Results of the neural networks on the test-set.

		Browning Score MAE		**Brown/Healthy Accuracy**	
		Per-Image	**Per-Apple**	**Per-Image**	**Per-Apple**
CT slices	Detection	0.33	0.20	96.7%	100%
	Prediction	0.19	0.00	90.6%	100%
Radiographs	Detection	0.43	0.40	85.7%	86.7%
	Prediction	0.63	0.60	78.6%	80.0%

Table 4. Results of the neural networks when they were applied to data from a different time point than what they were trained on.

		Browning Score MAE (Per-Image)	% Classified as Score 1	% Classified as Score 4
CT Slices	Detection network on just-out-of-CA-storage data	1.04	85.4%	0.0%
	Prediction network on after-shelf-life data	1.56	1.4%	87.8%
Radiographs	Detection network on just-out-of-CA-storage data	1.09	67.6%	0.0%
	Prediction network on after-shelf-life data	1.21	4.0%	60.4%

4.2. Workflow 2: Longitudinal XAI Heatmaps

Figure 2 shows an example of CT scans and simulated radiographs after image registration. The changes between the before-storage and later scans are shown as difference images. Some difference images show registration artifacts around the edges of the apple or the apple core, indicating some registration error. However, the registration error is very small compared to the features of interest. A visual inspection of the whole dataset showed registration errors in a similar range as visible in Figures 2 and 3.

Table 5 shows the (average) outputs of the neural networks on the different baselines. On healthy apples, the average output of the longitudinal baseline is close to one, so it can be considered a neutral baseline. For brown apples, outputs on the longitudinal baseline are slightly higher. Both the black and the zero baselines have outputs that are far outside of the 1–4 range on at least one network. In those cases, most of the intensity in the IG heatmaps explains the baseline rather than the input.

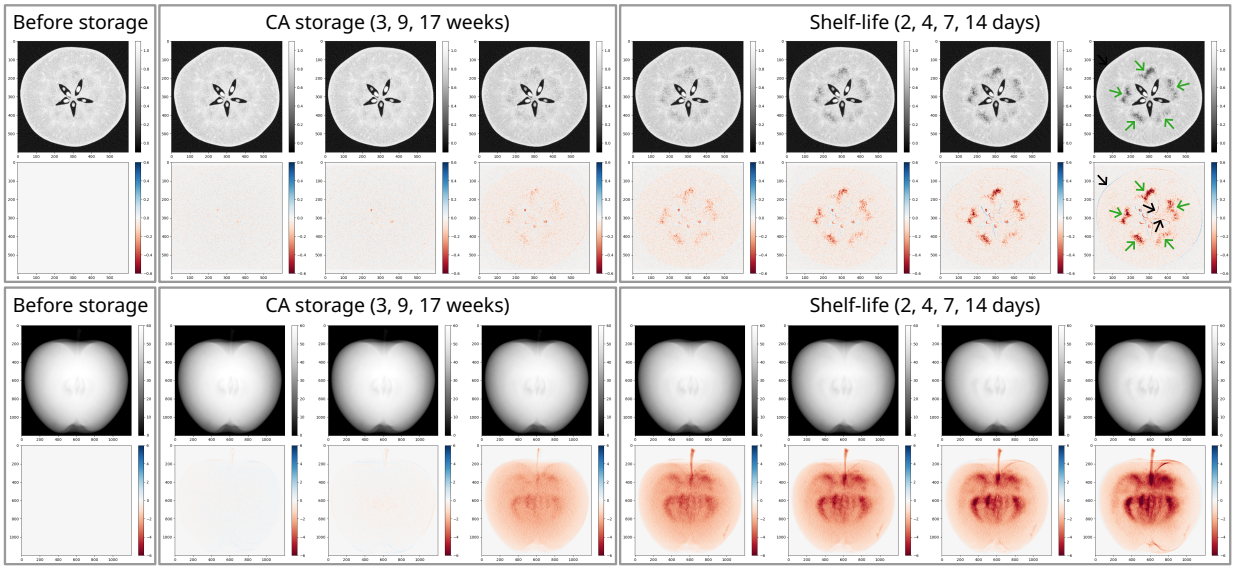


Figure 2. (**Top**) Central slice of registered CT scans over time, and difference images showing the difference from the before storage scan. (**Bottom**) Simulated radiographs of the same apple over time, and difference images showing the difference from the before-storage radiograph. The images show that browning develops both during CA storage and shelf life. The regions associated with browning are marked with green arrows in the last CT slice and difference image. When there is a registration error, it results in registration artifacts in the difference images around the edges of the apple or the apple core. These artifacts are marked with black arrows in the last CT difference image (you may only see them clearly if you zoom in). The registration errors are small compared to the browning regions.

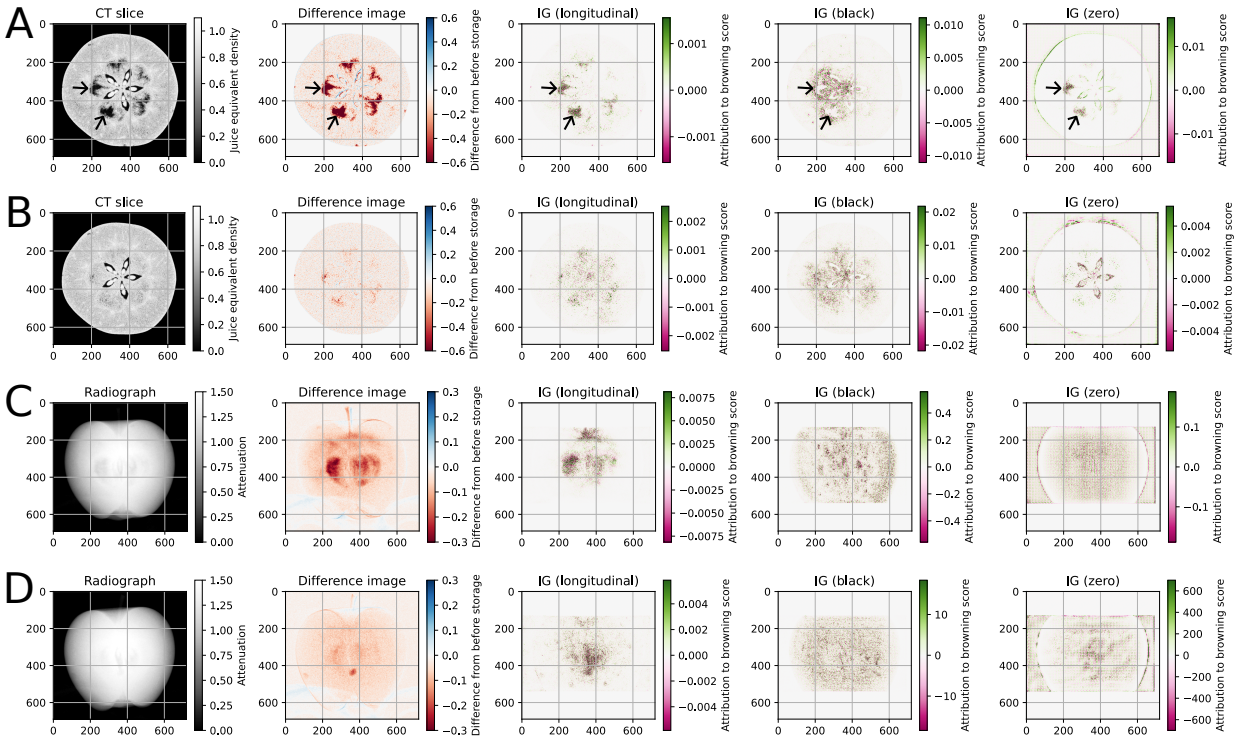


Figure 3. Examples of IG heatmaps with different baselines (longitudinal, black, or zero). (**A**) Result on the CT detection network. (**B**) Result on the CT prediction network. (**C**) Result on the radiograph detection network. (**D**) Result on the radiograph prediction network. All images were calculated on the same apple with browning score 4. IG with a black baseline is unable to highlight black regions, which explains why the two spots marked with arrows are not highlighted, even though they are strong indicators of browning. IG with a zero baseline highlights the background, even though this region is irrelevant in detecting browning.

Table 5. Neural network output for different baseline choices.

		Longitudinal on Healthy Apples	**Longitudinal on Brown Apples**	**Black**	**Zero**
CT Slices	Detection	1.00 ± 0.03	1.12 ± 0.23	1.12	−13.64
	Prediction	1.07 ± 0.07	1.55 ± 0.30	1.18	1.14
Radiographs	Detection	1.14 ± 0.20	1.54 ± 0.36	1.38	614.50
	Prediction	1.29 ± 0.05	1.29 ± 0.07	826.45	46,903.21

Figure 3 shows examples of IG heatmaps with the longitudinal, black, and zero baselines for all four networks. It shows that the heatmap with a black baseline (fourth column) on the CT detection network (first row) can not highlight the severely brown regions because they have the same intensity on the CT slice as on the baseline. Additional examples are included in Appendix A.

5. Discussion

5.1. Findings on the 'Braeburn' Browning Dataset

Our results in Table 3 align with existing research indicating that it is possible to detect browning after a shelf-life period from CT slices or radiographs [4] and to distinguish 4 classes of browning severity [23]. To the knowledge of the authors, there is no previous research on predicting browning from just-out-of-CA-storage CT slices or radiographs, making this the most promising finding. However, the sensitivity of developing browning in 'Braeburn' apples depends on the orchard [72] and weather conditions [29], but the

'Braeburn' browning dataset in this paper only contains data from a single orchard and a single growing season. Moreover, although the dataset contains a large number of CT slices and radiographs, they were acquired from a relatively small number of apples (Table 2). Therefore, while our our results on predicting internal browning are promising, experiments on additional data are required to verify their general applicability.

The worse MAE and accuracy on the radiography networks compared to the CT slice networks can be explained by the higher contrast of the CT slices. The CT slices also focused only on the center of the apple, which shows the most changes due to core browning. Combining the results from multiple images improved the results slightly. Earlier work [23] also suggested that a single CT slice is sufficient for detecting browning.

The detection and prediction networks presented here do not generalize to each other's data (Table 4). This is likely because the regions of the CT slice around the core that are associated with browning get darker during shelf-life (Figure 2), so the detection networks have been trained to detect darker spots than the prediction networks. Example images of the input data of the detection and prediction networks at each browning score are provided in Appendix A. These results suggest that, for grading browning, it is important to consider the moment of data acquisition relative to the progression of the disorder and the moment of consuming the apple.

All four networks showed higher average browning scores on the before-storage (longitudinal) baseline for apples that became brown after CA storage and a shelf-life period (Table 5). While this means the baseline is less neutral, it also suggests that the neural networks are using some information from the before-storage images to detect or predict browning, which can be considered a partial explanation. Moreover, these results suggest that it might be possible to use deep learning to identify apples before CA storage that would be predisposed to developing internal browning during (non-optimal) CA storage. This may be an interesting subject for future research.

5.2. Methodology of the Workflows

In this paper, the browning scores were modeled as a single continuous variable, so the neural networks used an image regression approach. It would also have been possible to model the browning scores as four separate classes, resulting in an image classification approach. Many XAI papers show examples on image classification but not on image regression [17,19,73]. However, for ordered classes such as the browning score, the regression approach used in this paper has the advantage that a single heatmap can explain the entire network. In the image classification case, a heatmap would be calculated for each class. Moreover, the outputs of image classification networks are constrained to sum to 1, so the network can only have a high output for one class. This would make it challenging to compare heatmaps of apples that were classified at different scores. There are also ordinal regression approaches [74,75] that maintain the discrete nature of the classes, but take into account their ordering. However, like image classification, these methods use multiple outputs, so the results on one image can not be explained by a single heatmap.

The main aim of this study was to demonstrate the two workflows. Therefore, we used only one network architecture (EfficientNetV2_s), and did not include an evaluation of data augmentation techniques and their hyperparameters. For use in industrial practice or for research on the optimal detection of specific disorders, we recommend evaluating multiple network architectures and data augmentation approaches.

The IG XAI method [17] was used in workflow 2 because it is commonly used and has been implemented in several explainable AI libraries, such as Captum and Saliency. The paper on IG also introduced the axiom-based approach of calculating XAI heatmaps. Recently, several similar methods for calculating heatmaps have been proposed, with the

aim of producing results that are less visually noisy. Several of these methods also require a baseline and satisfy the completeness axiom [19,44,47,48]. Therefore, it would be interesting future work to combine workflow 2 with these methods, as this may lead to less visually noisy heatmaps. The results from Table 5 only rely on evaluating the neural network at the baseline and input images, and on the completeness axiom, so they directly apply to these methods as well. Moreover, these methods also share the property that they are blind to areas that are the same between the input and the baseline.

5.3. Considerations for Applications in Practice

Acquisition time is a major constraint in real-world sorting systems. Commercial sorting systems have a throughput of roughly 10 apples per second, corresponding to a maximum acquisition time of 100 ms per sample. The radiographs used in this paper were acquired in 100 ms, so it is likely that a similar radiography-based system could be implemented in practice. The CT slices had a significantly longer acquisition time but higher accuracy, so the results on CT slices can serve as an indicator of the performance that is possible when there is no limit on acquisition time. The scanning setup used in this paper was not optimized for speed. Therefore, using optimized hardware is likely to yield similar performance with shorter acquisition times for both radiographs and CT slices. Research is ongoing to increase the throughput of CT scanners [76–78], so real-time CT scanning might become available in the future.

The prediction network in this paper predicted browning after a fixed two-week shelf-life period. However, by collecting a dataset with different shelf-life periods, it would be possible to train neural networks with multiple outputs, each for a different shelf-life period. This could be used to estimate best-before dates for each product individually. Such an approach could also be valuable in predicting ripeness in fruit with a narrow ripeness window, such as mangoes or avocados. Moreover, it might be possible to improve the estimations by providing the duration of CA storage as an additional input to the neural network.

Longitudinal XAI Heatmaps can also be applied to other imaging modalities, such as MRI or photography, and problems outside of the postharvest domain [54]. The main requirement is that the images are precisely aligned. This alignment may be achieved through image registration, as proposed in this paper, making the method more widely applicable. One example from medicine would be to extend the study of Wang et al. [79] on explaining networks for predicting a person's age from an MRI scan of their brain, which was done on the longitudinal Rotterdam study dataset. An image registration-based longitudinal baseline could be used to limit the explanation to changes occurring after a certain age. Moreover, it may reduce artifacts compared to other baselines, such as the black or zero baseline, which are also outside the data distribution for brain MRI images.

6. Conclusions

This paper presented two workflows for using longitudinal CT datasets in the development of X-ray-based postharvest disorder detection systems, and demonstrated them on a dataset of 'Braeburn' apples developing internal browning. The first workflow can be used to compare the detection accuracy at different time points within the development of the disorder, and to test whether networks generalize over time. On the 'Braeburn' browning dataset, the browning score could be detected after a shelf-life period or predicted from just-out-of-CA-storage data using deep learning. However, these neural networks did not generalize to each other's data, indicating that it is important to consider the time point of image acquisition relative to the development of a disorder. The second workflow can be used to generate XAI heatmaps that only explain changes over time. On

the 'Braeburn' browning dataset, this resulted in heatmaps that were not blind to regions of severe browning and did not exhibit severe activations in the background. While a more varied dataset is required to confirm our findings on the prediction of browning in 'Braeburn' apples, the results in this paper are a promising step towards developing real-world, explainable early detection systems for this disorder. Moreover, the workflows in this paper are sufficiently general to be applied to other 3D imaging modalities and to other gradually developing disorders, even outside the postharvest domain, which may benefit a wide range of future research.

Author Contributions: Conceptualization: D.E.S., R.M.W., R.S., R.v.L., T.v.L., and K.J.B.; Methodology: D.E.S. and R.M.W.; Software: D.E.S.; Validation, D.E.S.; Formal analysis: D.E.S.; Investigation: D.E.S. and R.M.W.; Resources: R.M.W. and R.S.; Data curation: D.E.S. and R.M.W.; Writing—original draft preparation: D.E.S.; Writing—review and editing: D.E.S., R.M.W., R.S., R.v.L., T.v.L., and K.J.B.; Visualization: D.E.S.; Supervision: R.S., R.v.L., T.v.L., and K.J.B.; Project administration: D.E.S., R.M.W., R.v.L., T.v.L., and K.J.B.; Funding acquisition, K.J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Dutch Research Council (NWO), grant number ENWSS.2018.003.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The large size of the dataset is the reason for not publishing the data alongside the paper. All computational methods were implemented in Python (version 3.11), and the code is publicly available on Github (https://github.com/D1rk123/longitudinal_ct_workflow, accessed on 6 July 2026).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Additional Example Images

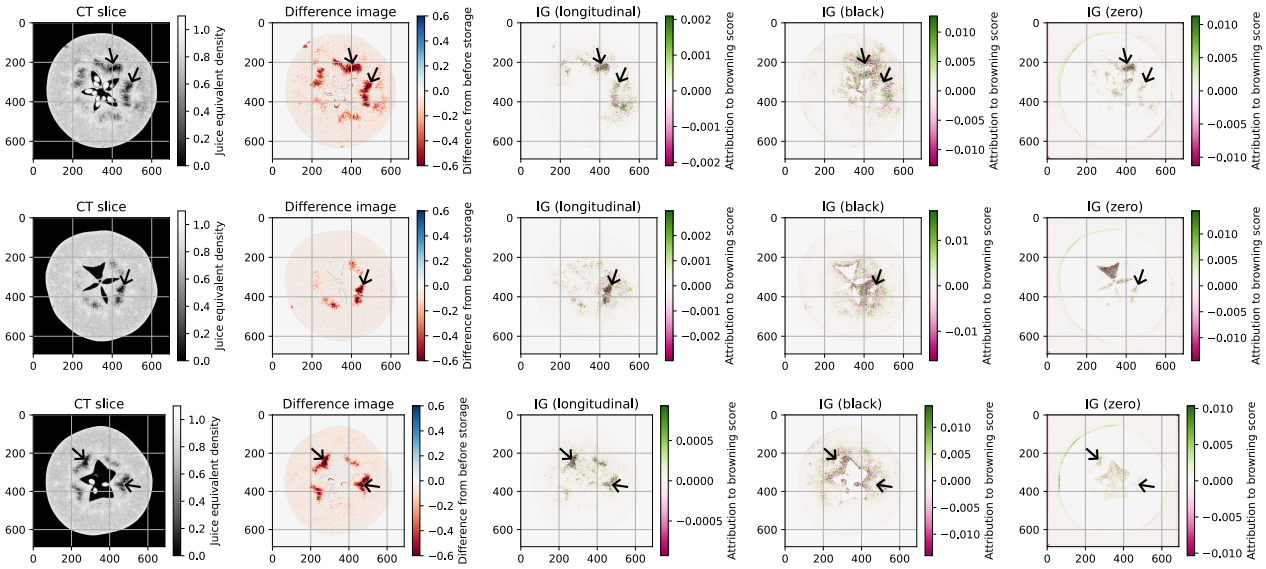


Figure A1. Additional examples of IG heatmaps with different baselines (longitudinal, black, or zero) on the CT detection network. IG with a black baseline is unable to highlight black regions, which explains why the spots marked with arrows are not highlighted, even though they are strong indicators of browning.

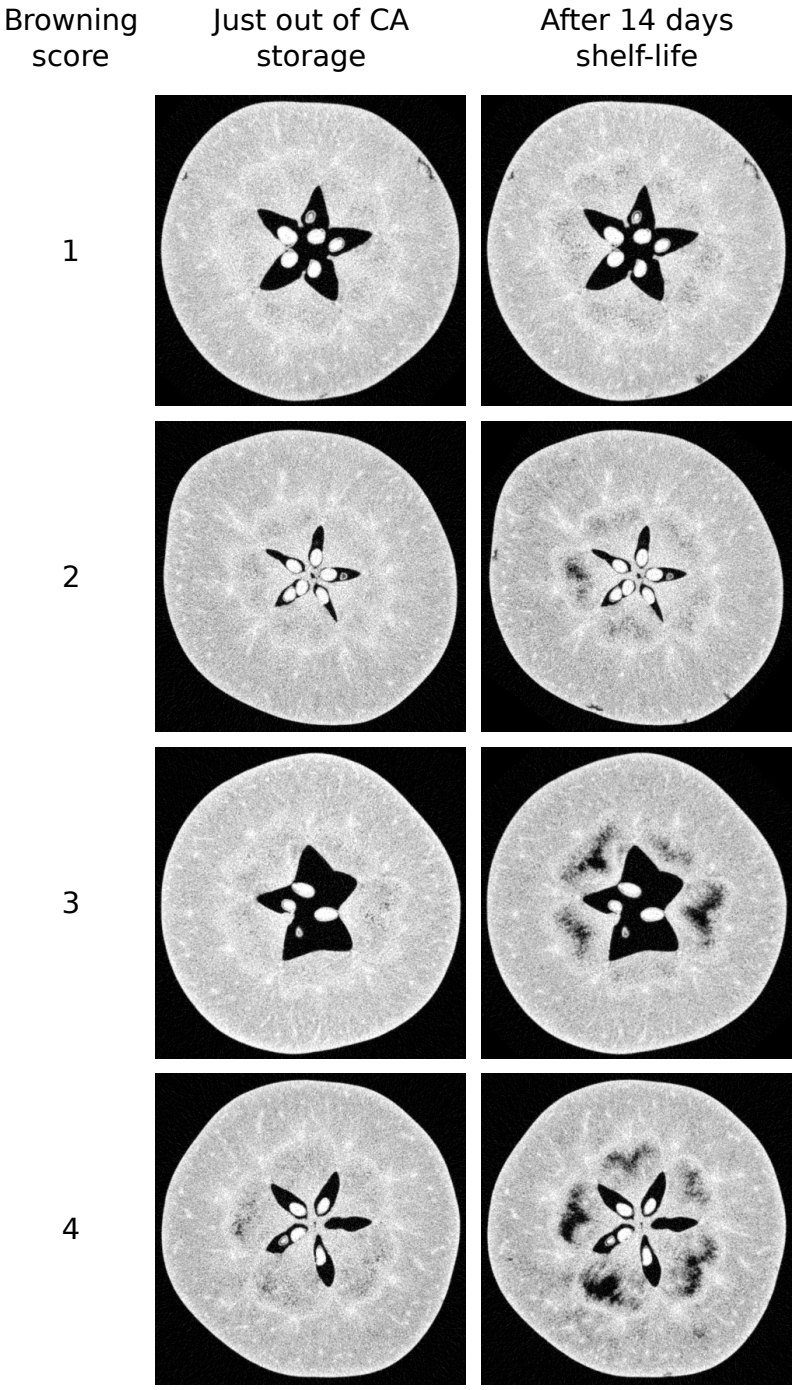


Figure A2. Example CT slices at different browning scores and scanning time points. The *Just out of CA storage* images on the left are used as input to the prediction networks, and the *After 14 days shelf-life* images on the right are used as input to the detection networks. The dark spots around the core that are associated with browning are a lot darker on the *After 14 days shelf-life* images than on the *Just out of CA storage* images. This difference explains why the detection and prediction networks did not generalize to each other's data (Table 4).

References

1. Gustavsson, J.; Cederberg, C.; Sonesson, U.; Van Otterdijk, R.; Meybeck, A. *Global Food Losses and Food Waste*; Technical report; Food and Agriculture Organization of the United Nations (FAO): Rome, Italy, 2011.
2. Nicolai, B.M.; Defraeye, T.; De Ketelaere, B.; Herremans, E.; Hertog, M.L.; Saeys, W.; Torricelli, A.; Vandendriessche, T.; Verboven, P. Nondestructive measurement of fruit and vegetable quality. *Annu. Rev. Food Sci. Technol.* **2014**, *5*, 285–312. [CrossRef] [PubMed]
3. Kotwaliwale, N.; Singh, K.; Kalne, A.; Jha, S.N.; Seth, N.; Kar, A. X-ray imaging methods for internal quality evaluation of agricultural produce. *J. Food Sci. Technol.* **2014**, *51*, 1–15. [CrossRef] [PubMed]

4. Tempelaere, A.; Van Doorselaer, L.; He, J.; Verboven, P.; Nicolai, B.M. BraeNet: Internal disorder detection in 'Braeburn' apple using X-ray imaging data. *Food Control* **2024**, *155*, 110092. [CrossRef]
5. He, J.; Belien, T.; Alhmedi, A.; Schrenk, A.; Bylemans, D.; Verboven, P.; Nicolai, B. Deep transfer learning for Codling moth damage detection in 'Conference' pear using X-ray radiography. *Food Control* **2025**, *183*, 111923. [CrossRef]
6. Yang, Z.; Zhang, J.; Li, Z.; Hu, N.; Qi, Z. Pears Internal Quality Inspection Based on X-Ray Imaging and Multi-Criteria Decision Fusion Model. *Agriculture* **2025**, *15*, 1315. [CrossRef]
7. Andriiashen, V.; van Liere, R.; van Leeuwen, T.; Batenburg, K.J. CT-based data generation for foreign object detection on a single X-ray projection. *Sci. Rep.* **2023**, *13*, 1881. [CrossRef] [PubMed]
8. Kiran, P.R.; Avinash, G.; Ray, M.; Nigam, S.; Parray, R.A. Deep learning models for detection and classification of spongy tissue disorder in mango using X-ray images. *J. Food Meas. Charact.* **2024**, *18*, 7806–7818. [CrossRef]
9. Yang, E.C.; Yang, M.M.; Liao, L.H.; Wu, W.Y.; Chen, T.W.; Chen, T.M.; Lin, T.T.; Jiang, J.A. 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. *Formos. Entomol.* **2006**, *26*, 171–186. [CrossRef]
10. Jarolmasjed, S.; Espinoza, C.Z.; Sankaran, S.; Khot, L.R. Postharvest bitter pit detection and progression evaluation in 'Honeycrisp'apples using computed tomography images. *Postharvest Biol. Technol.* **2016**, *118*, 35–42. [CrossRef]
11. Muziri, T.; Theron, K.; Cantre, D.; Wang, Z.; Verboven, P.; Nicolai, B.; Crouch, E. Microstructure analysis and detection of mealiness in 'Forelle'pear (*Pyrus communis* L.) by means of X-ray computed tomography. *Postharvest Biol. Technol.* **2016**, *120*, 145–156. [CrossRef]
12. Wood, R.M.; Schut, D.E.; Balk, P.A.; Trull, A.K.; Marcelis, L.F.; Schouten, R.E. Assessing the development of internal disorders in pome fruit with X-ray CT before, during and after controlled atmosphere storage and shelf life. *Food Control* **2025**, *168*, 110970. [CrossRef]
13. Van Dael, M.; Verboven, P.; Zanella, A.; Sijbers, J.; Nicolai, B. Combination of shape and X-ray inspection for apple internal quality control: In silico analysis of the methodology based on X-ray computed tomography. *Postharvest Biol. Technol.* **2019**, *148*, 218–227. [CrossRef]
14. Phuong, N.T.H.; Van, T.T.; Nkede, F.N.; Tanaka, F.; Tanaka, F. Preservation of strawberries using chitosan incorporated with lemongrass essential oil: An X-ray computed tomography analysis of the internal structure and quality parameters. *J. Food Eng.* **2024**, *361*, 111737. [CrossRef]
15. Phuong, N.T.H.; Tanaka, F.; Wardana, A.A.; Van, T.T.; Yan, X.; Nkede, F.N.; Tanaka, F. Persimmon preservation using edible coating of chitosan enriched with ginger oil and visualization of internal structure changes using X-ray computed tomography. *Int. J. Biol. Macromol.* **2024**, *262*, 130014. [CrossRef] [PubMed]
16. Tanaka, F.; Nashiro, K.; Obatake, W.; Tanaka, F.; Uchino, T. Observation and analysis of internal structure of cucumber fruit during storage using X-ray computed tomography. *Eng. Agric. Environ. Food* **2018**, *11*, 51–56. [CrossRef]
17. Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic attribution for deep networks. In *Proceedings of the International Conference on Machine Learning*; PMLR: Cambridge, MA, USA, 2017; pp. 3319–3328.
18. Lundstrom, D.; Razaviyayn, M. Four axiomatic characterizations of the integrated gradients attribution method. *J. Mach. Learn. Res.* **2025**, *26*, 1–31.
19. Kapishnikov, A.; Venugopalan, S.; Avci, B.; Wedin, B.; Terry, M.; Bolukbasi, T. Guided integrated gradients: An adaptive path method for removing noise. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2021; pp. 5050–5058. [CrossRef]
20. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2017; pp. 618–626. [CrossRef]
21. Bhati, D.; Amiruzzaman, M.; Zhao, Y.; Guercio, A.; Le, T. A survey of post-hoc xai methods from a visualization perspective: Challenges and opportunities. *IEEE Access* **2025**, *13*, 120785–120806. [CrossRef]
22. Kaur, H.; Nori, H.; Jenkins, S.; Caruana, R.; Wallach, H.; Wortman Vaughan, J. Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–14. [CrossRef]
23. Wood, R.M.; Schut, D.E.; Trull, A.K.; Marcelis, L.F.; Schouten, R.E. Detecting internal browning in apple tissue as determined by a single CT slice in intact fruit. *Postharvest Biol. Technol.* **2024**, *211*, 112802. [CrossRef]
24. Hendriksen, A.; Schut, D.; Palenstijn, W.J.; Viganò, N.; Kim, J.; Pelt, D.; van Leeuwen, T.; Batenburg, K.J. Tomosipo: Fast, Flexible, and Convenient 3D Tomography for Complex Scanning Geometries in Python. *Opt. Express* **2021**, *29*, 40494–40513. [CrossRef] [PubMed]
25. Sidhu, R.S.; Bound, S.A.; Swarts, N.D. Internal flesh browning in apple and its predisposing factors—A review. *Physiologia* **2023**, *3*, 145–172. [CrossRef]

26. Nicolas, J.J.; Richard-Forget, F.C.; Goupy, P.M.; Amiot, M.J.; Aubert, S.Y. Enzymatic browning reactions in apple and apple products. *Crit. Rev. Food Sci. Nutr.* **1994**, *34*, 109–157. [CrossRef] [PubMed]
27. Herremans, E.; Verboven, P.; Defraeye, T.; Rogge, S.; Ho, Q.T.; Hertog, M.L.; Verlinden, B.E.; Bongaers, E.; Wevers, M.; Nicolai, B.M. X-ray CT for quantitative food microstructure engineering: The apple case. *Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. At.* **2014**, *324*, 88–94. [CrossRef]
28. Schut, D.E.; Wood, R.M.; Trull, A.K.; Schouten, R.; van Liere, R.; van Leeuwen, T.; Batenburg, K.J. Joint 2D to 3D image registration workflow for comparing multiple slice photographs and CT scans of apple fruit with internal disorders. *Postharvest Biol. Technol.* **2024**, *211*, 112814. [CrossRef]
29. McCormick, R.; Biegert, K.; Streif, J. Occurrence of physiological browning disorders in stored 'Braeburn' apples as influenced by orchard and weather conditions. *Postharvest Biol. Technol.* **2021**, *177*, 111534. [CrossRef]
30. Lau, O. Effect of growing season, harvest maturity, waxing, low O_2 and elevated CO_2 on flesh browning disorders in 'Braeburn' apples. *Postharvest Biol. Technol.* **1998**, *14*, 131–141. [CrossRef]
31. Nugraha, B.; Verboven, P.; Verlinden, B.E.; Verreydt, C.; Boone, M.; Josipovic, I.; Nicolai, B.M. Gas exchange model using heterogeneous diffusivity to study internal browning in 'Conference' pear. *Postharvest Biol. Technol.* **2022**, *191*, 111985. [CrossRef]
32. Lammertyn, J.; Dresselaers, T.; Van Hecke, P.; Jancsók, P.; Wevers, M.; Nicolaı, B. Analysis of the time course of core breakdown in 'Conference' pears by means of MRI and X-ray CT. *Postharvest Biol. Technol.* **2003**, *29*, 19–28. [CrossRef]
33. Wang, J.; Lin, Y.; Li, Q.; Lu, Z.; Qian, J.; Dai, H.; Pi, F.; Liu, X.; He, Y. Non-destructive detection and grading of chilling injury-induced lignification of kiwifruit using X-ray computer tomography and machine learning. *Comput. Electron. Agric.* **2024**, *218*, 108658. [CrossRef]
34. Speir, R.A.; Haidekker, M.A. Onion postharvest quality assessment with X-ray computed tomography—A pilot study. *IEEE Instrum. Meas. Mag.* **2017**, *20*, 15–19. [CrossRef]
35. Karmoker, P.; Obatake, W.; Tanaka, F.; Tanaka, F. Visualization of porosity and thermal conductivity distributions of Japanese apricot and pear during storage using X-ray computed tomography. *Eng. Agric. Environ. Food* **2019**, *12*, 505–510. [CrossRef]
36. Rahman, M.M.; Joardder, M.U.; Karim, A. Non-destructive investigation of cellular level moisture distribution and morphological changes during drying of a plant-based food material. *Biosyst. Eng.* **2018**, *169*, 126–138. [CrossRef]
37. Prawiranto, K.; Defraeye, T.; Derome, D.; Bühlmann, A.; Hartmann, S.; Verboven, P.; Nicolai, B.; Carmeliet, J. Impact of drying methods on the changes of fruit microstructure unveiled by X-ray micro-computed tomography. *RSC Adv.* **2019**, *9*, 10606–10624. [CrossRef] [PubMed]
38. Zitova, B.; Flusser, J. Image registration methods: A survey. *Image Vis. Comput.* **2003**, *21*, 977–1000. [CrossRef]
39. Tempelaere, A.; De Ketelaere, B.; He, J.; Kalfas, I.; Pieters, M.; Saeys, W.; Van Belleghem, R.; Van Doorselaer, L.; Verboven, P.; Nicolai, B.M. An introduction to artificial intelligence in machine vision for postharvest detection of disorders in horticultural products. *Postharvest Biol. Technol.* **2023**, *206*, 112576. [CrossRef]
40. Matsui, T.; Sugimori, H.; Koseki, S.; Koyama, K. Automated detection of internal fruit rot in Hass avocado via deep learning-based semantic segmentation of X-ray images. *Postharvest Biol. Technol.* **2023**, *203*, 112390. [CrossRef]
41. Zhang, H.; Ning, X.; Pu, H.; Ji, S. A novel approach for the non-destructive detection of shriveling degrees in walnuts using improved YOLOv5n based on X-ray images. *Postharvest Biol. Technol.* **2024**, *214*, 113007. [CrossRef]
42. Van De Looverbosch, T.; He, J.; Tempelaere, A.; Kelchtermans, K.; Verboven, P.; Tuytelaars, T.; Sijbers, J.; Nicolai, B. Inline nondestructive internal disorder detection in pear fruit using explainable deep anomaly detection on X-ray images. *Comput. Electron. Agric.* **2022**, *197*, 106962. [CrossRef]
43. Simonyan, K.; Vedaldi, A.; Zisserman, A. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv* **2013**, arXiv:1312.6034. [CrossRef]
44. Zhang, B.; Zheng, W.; Zhou, J.; Lu, J. Path Choice Matters for Clear Attributions in Path Methods. In *Proceedings of the International Conference on Learning Representations (ICLR)*; PMLR: Cambridge, MA, USA, 2024.
45. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2016; pp. 2921–2929. [CrossRef]
46. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; IEEE: New York, NY, USA, 2016; pp. 1135–1144. [CrossRef]
47. Sundararajan, M.; Najmi, A. The many Shapley values for model explanation. In *Proceedings of the International Conference on Machine Learning*; PMLR: Cambridge, MA, USA, 2020; pp. 9269–9278.
48. Schut, D.E.; van Liere, R.; van Leeuwen, T. Monte Carlo Multi-Feature Baseline Shapley (MMBS): An axiomatic attribution method for fine-grained explanations of image classification networks. *Trans. Mach. Learn. Res.* **2026**. Available online: https://openreview.net/forum?id=LLFIcr7zWh (accessed on 26 May 2026).
49. Smilkov, D.; Thorat, N.; Kim, B.; Viégas, F.; Wattenberg, M. Smoothgrad: Removing noise by adding noise. *arXiv* **2017**, arXiv:1706.03825. [CrossRef]

50. Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.R.; Samek, W. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLoS ONE* **2015**, *10*, e0130140. [CrossRef] [PubMed]
51. Achtibat, R.; Dreyer, M.; Eisenbraun, I.; Bosse, S.; Wiegand, T.; Samek, W.; Lapuschkin, S. From attribution maps to human-understandable explanations through concept relevance propagation. *Nat. Mach. Intell.* **2023**, *5*, 1006–1019. [CrossRef]
52. Sturmfels, P.; Lundberg, S.; Lee, S.I. Visualizing the impact of feature attribution baselines. *Distill* **2020**, *5*, e22. [CrossRef]
53. Sundararajan, M.; Taly, A. A Note about: Local Explanation Methods for Deep Neural Networks lack Sensitivity to Parameter Values. *arXiv* **2018**, arXiv:1806.04205. [CrossRef]
54. Mamalakis, A.; Barnes, E.A.; Ebert-Uphoff, I. Carefully choose the baseline: Lessons learned from applying XAI attribution methods for regression tasks in geoscience. *Artif. Intell. Earth Syst.* **2023**, *2*, e220058. [CrossRef]
55. Raju, A.S.N.; Sujatha, G.; Gatla, R.K.; Ankalaki, S. SpinachXAI-Rec: A multi-stage explainable AI framework for spinach freshness classification and consumer recommendation. *Sci. Rep.* **2025**, *15*, 35853. [CrossRef] [PubMed]
56. Srinivasan, S.; Somasundharam, L.; Rajendran, S.; Singh, V.P.; Mathivanan, S.K.; Moorthy, U. DBA-ViNet: An effective deep learning framework for fruit disease detection and classification using explainable AI. *BMC Plant Biol.* **2025**, *25*, 965. [CrossRef] [PubMed]
57. Lee, I.H.; Li, Z.; Ma, L. Explainable AI and mobile imaging for non-destructive avocado ripeness and internal quality assessment to reduce food waste. *Curr. Res. Food Sci.* **2025**, *11*, 101196. [CrossRef] [PubMed]
58. Feldkamp, L.A.; Davis, L.C.; Kress, J.W. Practical cone-beam algorithm. *J. Opt. Soc. Am. A* **1984**, *1*, 612–619. [CrossRef]
59. Herman, G.T. Correction for beam hardening in computed tomography. *Phys. Med. Biol.* **1979**, *24*, 81. [CrossRef] [PubMed]
60. Schoonhoven, R.; Skorikov, A.; Palenstijn, W.J.; Pelt, D.M.; Hendriksen, A.A.; Batenburg, K.J. How auto-differentiation can improve CT workflows: Classical algorithms in a modern framework. *Opt. Express* **2024**, *32*, 9019–9041. [CrossRef] [PubMed]
61. Tan, M.; Le, Q. Efficientnetv2: Smaller models and faster training. In *Proceedings of the International Conference on Machine Learning*; PMLR: Cambridge, MA, USA, 2021; pp. 10096–10106.
62. Marcel, S.; Rodriguez, Y. Torchvision the machine-vision package of torch. In *Proceedings of the 18th ACM International Conference on Multimedia*; Association for Computing Machinery: New York, NY, USA, 2010; pp. 1485–1488. [CrossRef]
63. Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albumentations: Fast and Flexible Image Augmentations. *Information* **2020**, *11*, 125. [CrossRef]
64. Simard, P.Y.; Steinkraus, D.; Platt, J.C. Best practices for convolutional neural networks applied to visual document analysis. In *Seventh International Conference on Document Analysis and Recognition*; IEEE: New York, NY, USA, 2003; Volume 3. [CrossRef]
65. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic differentiation in PyTorch. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
66. Falcon, W.; The PyTorch Lightning Team. *PyTorch Lightning*, version 2.2.1; Zenodo: Geneva, Switzerland, 2024. [CrossRef]
67. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101. [CrossRef]
68. Lowekamp, B.C.; Chen, D.T.; Ibáñez, L.; Blezek, D. The design of SimpleITK. *Front. Neuroinform.* **2013**, *7*, 45. [CrossRef] [PubMed]
69. Avants, B.B.; Tustison, N.J.; Stauffer, M.; Song, G.; Wu, B.; Gee, J.C. The Insight ToolKit image registration framework. *Front. Neuroinform.* **2014**, *8*, 44. [CrossRef] [PubMed]
70. Liu, D.C.; Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. Program.* **1989**, *45*, 503–528. [CrossRef]
71. Kokhlikyan, N.; Miglani, V.; Martin, M.; Wang, E.; Alsallakh, B.; Reynolds, J.; Melnikov, A.; Kliushkina, N.; Araya, C.; Yan, S.; et al. Captum: A unified and generic model interpretability library for pytorch. *arXiv* **2020**, arXiv:2009.07896. [CrossRef]
72. Neuwald, D.A.; Sestari, I.; Kittemann, D.; Streif, J.; Weber, A.; Brackmann, A. 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. *Erwerbs-Obstbau* **2014**, *56*, 35–41. [CrossRef]
73. Zhang, Y.; Liu, Q.; Chen, J.; Sun, C.; Lin, S.; Cao, H.; Xiao, Z.; Huang, M. Developing non-invasive 3D quantificational imaging for intelligent coconut analysis system with X-ray. *Plant Methods* **2023**, *19*, 24. [CrossRef] [PubMed]
74. Cheng, J.; Wang, Z.; Pollastri, G. A neural network approach to ordinal regression. In *Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*; IEEE: New York, NY, USA, 2008; pp. 1279–1284.
75. Diaz, R.; Marathe, A. Soft labels for ordinal regression. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2019; pp. 4738–4747.
76. Morton, E.; Mann, K.; Berman, A.; Knaup, M.; Kachelrieß, M. Ultrafast 3D reconstruction for X-ray real-time tomography (RTT). In *Proceedings of the 2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*; IEEE: New York, NY, USA, 2009; pp. 4077–4080. [CrossRef]
77. Zwanenburg, E.; Williams, M.; Warnett, J.M. Review of high-speed imaging with lab-based x-ray computed tomography. *Meas. Sci. Technol.* **2022**, *33*, 012003. [CrossRef]

78. De Schryver, T.; Dhaene, J.; Dierick, M.; Boone, M.N.; Janssens, E.; Sijbers, J.; van Dael, M.; Verboven, P.; Nicolai, B.; Van Hoorebeke, L. In-line NDT with X-Ray CT combining sample rotation and translation. *NDT E Int.* **2016**, *84*, 89–98. [CrossRef]
79. Wang, J.; Knol, M.J.; Tiulpin, A.; Dubost, F.; de Bruijne, M.; Vernooij, M.W.; Adams, H.H.; Ikram, M.A.; Niessen, W.J.; Roshchupkin, G.V. Gray matter age prediction as a biomarker for risk of dementia. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 21213–21218. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.