Deep Learning for Apple Fruit Quality Inspection using X-Ray Imaging

Astrid Tempelaere
KU Leuven, Belgium
astrid.tempelaere@kuleuven.be

Leen Van Doorselaer
leen.vandoorselaer@kuleuven.be

Jiaqi He
jiaqi.he@kuleuven.be

Pieter Verboven
pieter.verboven@kuleuven.be

Tinne Tuytelaars
tinne.tuytelaars@kuleuven.be

Bart Nicolai
bart.nicolai@kuleuven.be

Abstract

Apples are widely consumed worldwide, but the quality of the fruit flesh might deteriorate during storage, resulting in brown tissue formation. X-ray radiography has emerged as a non-destructive method for quickly detecting internal quality problems. This method provides X-ray imaging data that should be processed in an accurate and efficient way. In this paper, we investigate the classification of healthy and defect apples from different orchards and storage conditions using deep learning. The aim of the study was to select a robust and efficient deep learning network that can be used on an X-ray sorting system in a practical setting in the agrifood industry. To this end, the models were evaluated not only in terms of performance but also computational cost. As biological variability is inherent to agrifood problems, we strongly focused on generalizability of the models by using multiple test sets with apples from another orchard and stored under different conditions. The best model had the GoogLeNet architecture, reaching an accuracy of respectively 100 (0)% on a first test set with apples from another orchard, and 82 (8)% on a second test set stored at other conditions. The comparative study provides valuable insights for improving robust and efficient detection algorithms and implementing X-ray technology in the agrifood industry. The proposed technology can be extended to other fruit and vegetables that also suffer from internal quality problems.

1. Introduction

To ensure a year-round supply of fresh produce, fruit and vegetables are typically stored under controlled atmosphere conditions. However, quality deterioration can still occur due to suboptimal conditions or the inherent susceptibility of certain products. Apples, a highly consumed commodity worldwide, are prone to internal quality deterioration, often appearing as browning of the fruit flesh [12, 17, 34]. In today's agrifood industry, only a small sample size of apple fruit per batch is evaluated by cutting open the fruit, but the entire batch is discarded in case of a negative evaluation. This leads to large food losses of a significant fraction of healthy products in the batch. Postharvest losses due to quality decay, often caused by disorders, affect 5 to 25% of total horticultural crop production [35].

Recently, X-ray radiography has been proposed as a non-destructive and fast technique to detect the internal quality in fruit and vegetables, as demonstrated for apple [26, 27, 28], pear [38, 40, 42], citrus [10, 37], seeds and grains [1, 2, 4], and many other commodities [7, 9, 29].

2. Related work

Over the years, image processing has developed from visual inspection to machine learning. Last years, deep learning for machine vision tasks such as classification has achieved a great success. In 1998, Lecun *et al.* [18] introduced the first convolutional neural network (CNN) for classification. In the following years, various state-of-the-art CNNs have been proposed, such as AlexNet [16], GoogLeNet [31], VGG [30], ResNet [11], EfficientNet [32], and ResNeXt [41]. While CNNs have been the standard in vision tasks for many years, natural language processing has developed differently, with Transformer architectures becoming the norm. Recently, vision and language research has converged, aiming to apply Transformers on vision tasks, such as the Vision Transformer (ViT) [8]. Inspired by this, Liu *et al.* [21] started from ResNet and gradually incorporated Transformer elements to improve performance, resulting in the ConvNeXt model.

In food quality inspection, prior work on deep learning used simulated X-ray radiographs for pear and a ResNet classifier [11, 40]. Also grain and seed quality [4, 24] was inspected by AlexNet [16], VGG [30], ResNet [11], Inception-ResNet [31], Xception [3], and MobileNetV2 [25]. Besides these CNNs, ViT has also found its way towards food imaging, such as the detection of diseased kiwi

[20] and cassava leaves [36], and the classification of diverse types of vegetables [19].

The earlier agricultural studies have often focused on one or a few specific network architectures, evaluated on a small test set of foods from the same batch. In addition, models have sometimes been developed and tested on simulated X-ray data, but not further evaluated for real X-ray images from an industrial machine [40]. Hence, to obtain a deep learning algorithm that is applicable in the agrifood industry, the most efficient and effective architecture should be identified as well as it should be thoroughly evaluated on more diverse test sets that cover the biological variability inherent to nature. Another hurdle is that the datasets in agrifood studies typically consist of just a hundred of samples. This is in contrast to the state-of-the-art models that are designed for large public datasets, such as ImageNet [6]. In addition, such public datasets consist of RGB images, while X-ray radiographs are one-channel data.

In this work, we leverage the history of network architectures for the classification of healthy and defect apples based on X-ray images. We apply basic state-of-the-art deep learning models. Thereto, we focus on the model performance for multiple independent test sets and its computational workload. Our findings and discussion aim to encourage rethinking the importance of efficient model design and convolutional operations in computer vision. The ultimate goal of this study is to select an efficient and robust deep learning model for accurate detection of internal quality of apples from different orchards and storage conditions, that can rapidly run on an available industrial X-ray sorting machine. In general, the aimed throughput of such a device is ten samples per second to meet the speed of commercial sorting lines for external quality attributes, such as color and shape. Such an X-ray scanner with a conveyor belt is already commercially available. It can even be simply combined with a small embedded GPU for inference. The X-ray technology has been approved to be safe on food for human consumption, with today's application in the food industry mainly laying in foreign object detection [5, 39]. The main contributions of this work are as follows.

- We collected and curated three datasets of radiography images with binary labels for detecting internal disorders in apple fruit. The datasets include fruit from different orchards and storage conditions and can serve as a benchmark for the research community. The images are manually labeled as 'healthy' or 'defect' based on RGB images of cut-open fruit. The datasets will be made available upon reasonable request.
- Ten state-of-the-art pretrained models were trained using ten-fold stratified cross-validation: AlexNet, GoogLeNet, VGG16, ResNet18, ResNet50, EfficientNet-B0, EfficientNet-B1, ResNeXt50, ViT, and ConvNeXt.
- The model performance and generalizability was evaluated on two independent datasets with apples from another orchard and storage condition.
- The ten models were evaluated in terms of accuracy, recall, and precision, in addition to the computational requirements. The latter aspect is essential for developing an efficient and fast sorting system in the agrifood industry.

3. Datasets

Apple fruit (cultivar Braeburn) was harvested from different orchards and stored under diverse controlled atmosphere conditions. Table 1 gives an overview of the number of samples per dataset, later used as training and test data for classification. All fruit were harvested in the late picking window of 2021. The fruit underwent a cooling period of 21 days before applying the optimal conditions, consisting of regular air or 0.7% CO_2 and 2.5% O_2. Other fruit was immediately placed under controlled atmosphere for the disorder conditions, consisting of anoxia (100% N_2) for 4 weeks or hypoxia (10% CO_2 and 1% O_2) for 3.5 months. Prior to X-ray imaging, all fruit were put in shelf-life condition (18°C) for three days.

From all fruit in Table 1, X-ray radiographs were collected using a prototype industrial line scanner with conveyor belt (InnospeXion), available in the lab (Figure 1). Four radiographs per fruit were collected by turning the fruit around its longitudinal axis for 90 degrees. The X-ray source operated at 35 kV and 5000 mA. The X-ray detector had a resolution of 8160 x 256 pixels, a pixel size of 27 µm, and line rate of 2000 Hz. The pixels in the acquired data underwent a binning process with a factor of five, resulting in an image comprising 1632 x 2000 pixels. The speed of the conveyor belt was set at 0.27 m/s. Each apple was placed in a Styrofoam sample folder to fix its position during scanning. In the preprocessing step, the X-ray images were subjected to darkfield and flatfield correction. The sample holder was removed by thresholding the image, and a gamma correction of 0.5 was applied to the image data. The object of interest was cropped out to 1632 x 1632 pixels and the images were resized to 224 x 224 pixels.

After X-ray imaging, all apple samples were cut open to visually inspect the internal quality. All apples under optimal conditions were unaffected, while all these under disorder conditions developed internal browning. These observations were used as ground-truth labels ('healthy' or 'defect') for each individual apple.

Dataset	Optimal condition	Disorder condition	Orchard
Training set	n=30, regular air, 3°C	n=30, 100% N_2, 3°C	50.82°N, 5.28°E
Test set (1)	n=14, regular air, 3°C	n=30, 100% N_2, 3°C	50.79°N, 5.37°E
Test set (2)	n=107, 0.7% CO_2, 2.5% O_2, 1°C	n=168, 10% CO_2, 1% O_2, 1°C	50.82°N, 4.80°E

Table 1. Overview of the collected datasets (n, number of samples)

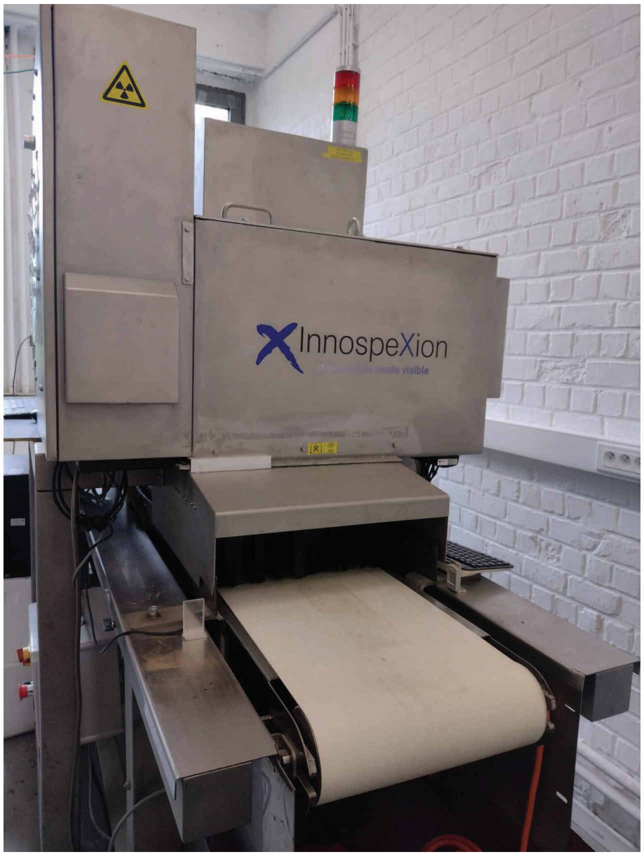

Figure 1. Prototype industrial line scanner with conveyor belt (InnospeXion), available in the lab. Device used for collecting X-ray imaging datasets.

4. Method

4.1. Approach

In this work, multiple state-of-the-art classification networks were applied to detect internal disorders in apple fruit based on X-ray imaging data. To this end, models were trained on a balanced training set consisting of 30 healthy and 30 disordered apples. From both labels, 27 (90%) and 3 (10%) samples were respectively used for model training and validation. Data augmentation including image flipping (horizontal, vertical, horizontal and vertical) and random affine transformation resulted in 1080 and 120 images for training and validation, respectively. Ten state-of-the-art ImageNet-pretrained networks were fine-tuned on this specific dataset. The models were trained in ten-fold stratified cross-validation, using a binary cross-entropy loss and Adam [14] and AdamW [22] optimizer ($ß_1$=0.9, $ß_2$=0.999) for the CNNs and Transformer-inspired (ViT, ConvNeXt) models, respectively, as in the original works. The batch size was kept at 256 and the initial learning rate (LR) was optimized during the network design (10^{-3}, 10^{-4}, 10^{-5}), and gradually decreased by factor 0.1 if no improvement was seen for the training loss for three epochs. We applied early stopping for all the architectures to avoid overfitting as evaluated by the accuracy on the validation set. The input dimensions at the first convolutional layer of each network architecture were adapted to 224 x 224 and the output channels of the last layer was set to one, followed by a sigmoid function to perform the binary classification (threshold at 0.5). Finally, ten models were obtained that were used for inference by providing them X-ray imaging data from the two independent test sets.

4.2. Network architectures

Ten pretrained classification models were fine-tuned using the PyTorch library [23] and an RTX A6000 Nvidia GPU.

- AlexNet [16]. The implementation provided in PyTorch was slightly different from the original one, and is based on [15]. LR = 10^{-4}. Epochs = 70.

- GoogLeNet [31]. The first version, i.e., Inception v1, was used. As the original PyTorch model only accepts three-channel input data, the model was preceded by a two-dimensional convolutional layer with a 1 x 1 convolutional operation to extend the one-channel input data to three channels. LR = 10^{-3}. Epochs = 50.

- VGG16 [30]. The VGG16 model with batch normalization was used. LR = 10^{-4}. Epochs = 50.

- ResNet [11]. A ResNet18 and ResNet50 model was used. LR = 10^{-3}. Epochs = 50.

- EfficientNet [32]. The baseline EfficientNet-B0 network was used as well as the scaled model EfficientNet-B1. LR = 10^{-3}. Epochs = 50.

- ResNeXt50 [41]. The version with a 32 x 4d template was used. LR = 10^{-3}. Epochs = 50.

- ViT [8]. The base version of this model with 16 x 16 input patch size, i.e., ViT-B/16, was used. AdamW optimizer. LR = 10^{-5}. Epochs = 50.

- ConvNeXt [21]. The tiny version of this model, i.e., ConvNeXt-T, was used. AdamW optimizer. LR = $5*10^{-5}$. Epochs = 50.

4.3. Model performance

Model performance on the two independent test sets (Table 1) was evaluated against ground-truth labels of 'healthy' and 'defect' using a confusion matrix that presents the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) classifications. Additional metrics, including the accuracy (1), precision (2), and recall (3) were calculated from these results. These metrics were calculated for all models obtained via ten-fold cross-validation and reported as median (interquartile range) as the datapoints were often not normally distributed.

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{1}$$

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

Besides the accuracy rate, the model complexity was analyzed by counting the total amount of learnable parameters and collecting the size of the parameter file in terms of megabyte (MB) for the considered models. In addition, the computational cost of each deep learning model was calculated via Giga Floating-Point Operations per Second (G-FLOPS). The information is useful for gaining insights into the powerful hardware required, such as GPU memory, for each model.

5. Results & discussion

5.1. Apples with different disorders

X-ray images were collected from the apples and these were labeled by cutting open the fruit. Figure 2A-C shows an example for disordered cut-open apples for each available dataset (Table 1) together with an X-ray radiograph. These images were compared to the data from a healthy apple in Fig. 2D. The N_2 condition for the training set and test set (1) resulted in internal defects appearing as dark brown speckles in the fruit flesh (Fig. 2A, B). On the other hand, the CO_2 condition caused large patches of browning and external injury symptoms visible just beneath the skin (Fig. 2C). This is in contrast to the fruit stored under air/optimal conditions, which all remained unaffected (Fig. 2D). The radiographs in Fig. 2 were adapted via Contrast Limited Adaptive Histogram Equalization (clip limit 5, tile grid 10) to better visualize the disorders. Some heterogeneities were observed in the radiographic images from the disordered fruit (Fig. 2A-C) compared to the radiograph from the healthy apple (Fig. 2D). However, relating these observations to the specific browning in the cut-open sample remained difficult. The other appearance of browning disorder in the fruit flesh stored under N_2 or CO_2 could also not be clearly observed in the radiographs, although an earlier study with X-ray computed tomography that resulted in 3D imaging data demonstrated that the large browning patches under CO_2 had a higher density compared to the healthy fruit tissue, while the brown speckles under N_2 had a lower density [33].

5.2. Deep learning architectures

During the last decade, advanced deep learning architectures have been designed with constantly improving accuracy on the ImageNet dataset. However, successful models could also be very large and therefore difficult to implement in a practical application. This work represents the history of state-of-the-art deep learning classifiers, and started from exploring some typical classification models in terms of size, top-1, and top-5 accuracy on the ImageNet-1K dataset. Table 2 depicts that the accuracy of these classification networks has gradually increased over the past years.

5.3. Generalizability

The success rate in deep learning is typically highly dependent on the network architecture, specific dataset, and classification task. For instance, some models may need much data or be more difficult to stabilize during training, such as Transformer-based models [8]. In this work, we applied the history of ImageNet-pretrained state-of-the-art classification architectures on 'healthy' vs 'defect' classification for X-ray images from 'Braeburn' apples. We focused on the generalizability of the models to diverse test sets, as well as the computational requirements for the different network architectures.

Deep learning models were trained on the training set consisting of healthy samples and defect ones due to N_2 storage (Table 1). A total of ten architectures were evaluated on two test sets using the accuracy (eq. 1), precision (eq. 2), and recall (eq. 3) as evaluation metrics, with results presented in Table 3. The numbers show that all models had a considerable higher performance on the first test set which consisted of apples stored under N_2, like the training set, but coming from another orchard. In contrast, the apples in the second test set were stored under CO_2. As stated above, the storage conditions led to a different appearance of disorders in the fruit flesh (Fig. 2). The models were thus trained on disorders arising from N_2, but also evaluated for their detection ability of CO_2 induced disorders.

From Table 3, it is clear that AlexNet, together with ViT

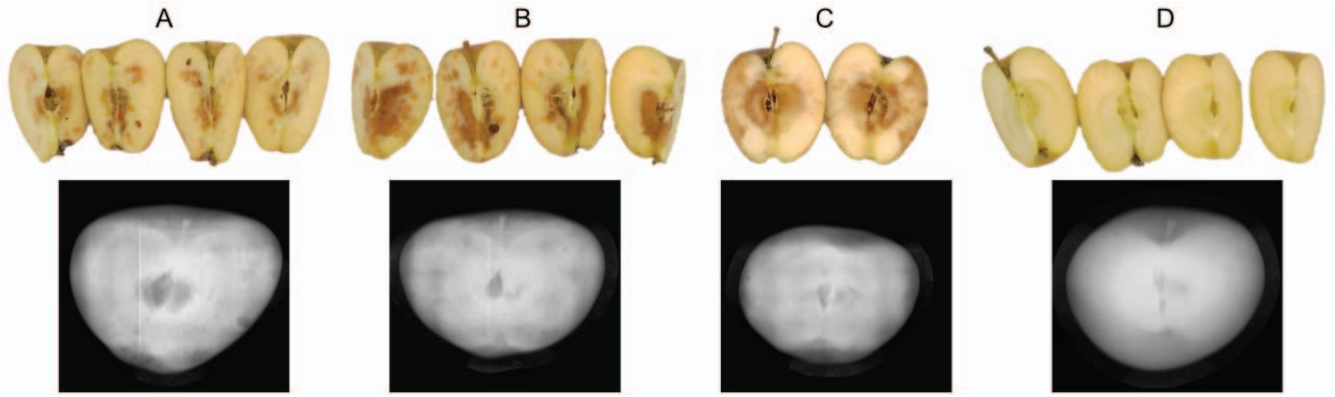

Figure 2. Image of a cut-open apple with X-ray radiograph. Disordered samples from A Training set, B Test set (1), and C Test set (2). Healthy apple from D Training set.

	AlexNet	GoogLeNet	VGG16	ResNet18	ResNet50	EffNet-B0	EffNet-B1	ResNeXt50	ViT	ConvNeXt
Year	2012	2013	2015	2015	2015	2016	2016	2017	2020	2022
Size (MB)	233.1	49.7	527.9	44.7	97.8	20.5	30.1	95.8	330.3	109.1
Top-1 acc. (%)	56.5	69.8	72.4	69.8	76.1	77.7	78.6	77.6	81.1	82.5
Top-5 acc. (%)	79.1	89.5	91.5	89.1	92.9	93.5	94.2	93.7	95.3	96.1

Table 2. Overview of the ten pretrained deep learning architectures used in this research and the top-1 and top-5 accuracy on the ImageNet-1K dataset.

and ConvNeXt had a considerably lower performance than all other models. To clarify the differences between the models, the network architectures were considered in detail.

AlexNet was one of the earliest CNNs [16], having a sequential architecture with five convolutional layers, and obtained an accuracy of 62 (2) and 50 (3)% on the two test sets. Following research extended the capabilities of the model in later CNNs. In 2014, Szegedy *et al*. [31] introduced the inception module, which performs convolutions using different sizes of filters on the same level, resulting in the GoogLeNet model. This inception module led to substantial improvement in performance of 100 (0) and 82 (8)% on the test sets. The kernels of size 1, 3, and 5 in a single inception module are thus more effective than using individual kernels in the convolutional layers of AlexNet. Further design of CNNs, with deeper models, residual training, and scalable modules in the VGG [30], ResNet [11], and EfficientNet [32] families, all resulted in similar performances. On the other hand, drawing conclusions for the performance on test set (2) is difficult because of the high interquartile range, calculated from the inference by the ten models trained for each architecture and train-validation split. The reason for this inferior performance for all models on test set (2) might be found in the data itself. As discussed above, the appearance of the CO_2 related disorders is different from the disorders due to N_2 storage. Hence, the extracted features for the latter disorders, on which the models are trained, are thus not sufficient to detect CO_2 related disorders. Generalizability is a common issue in agricultural vision tasks. For instance, Kamal *et al*. [13] reached a performance of 99.5% by using VGG on a public plant leaves dataset, but decreased to 33.3% for in real-world detection.

Besides CNNs, also upcoming Transformer-based models were considered. However, a ViT typically needs a large dataset and might be difficult to stabilize [8]. Recently, the ResNet50 model has been reintroduced as ConvNeXt using similar training procedures as ViT [21]. However, both ViT and ConvNeXt resulted in an inferior performance, with models having diverging accuracy according to the specific training-validation set folds as demonstrated by the high interquartile range.

The results are somewhat different from the accuracy on ImageNet. To explain, we notice from Table 2 that our most accurate models, i.e., GoogLeNet, VGG16, ResNet, EfficientNet, and ResNeXt, were outperformed by ViT and ConvNeXt for the ImageNet dataset. We fine-tuned all these pretrained models for our specific classification task, but it gave inferior results for the two latter architectures. A possible reason might be the limited number of training images used in our work.

Apart from the accuracy, the recall and precision were also provided. The first metric is important to circumvent that disordered fruit reaches the consumer. On the other hand, not too much healthy fruit may be regarded as disordered, as represented by a low precision, resulting in elevated food loss. Comparing the two test sets in more detail, a very high recall was observed for all models on test set

	AlexNet	GoogLeNet	VGG16	ResNet18	ResNet50	EffNet-B0	EffNet-B1	ResNeXt50	ViT	ConvNeXt
Test set (1)										
Acc. (%)	62 (2)	**100 (0)**	100 (0.7)	99.1 (0.7)	100 (0.9)	99.6 (0.9)	100 (0.6)	**100 (0)**	73 (7)	74 (12)
Recall (%)	96 (4)	**100 (0)**	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	**100 (0)**	99 (1)	100 (2)
Precision (%)	57 (1)	**100 (0)**	100 (1)	98 (1)	100 (2)	99 (2)	100 (1)	**100 (0)**	65 (6)	66 (8)
Test set (2)										
Acc. (%)	50 (3)	82 (8)	80 (13)	**85 (14)**	**85 (11)**	81 (4)	80 (9)	80 (9)	56 (10)	74 (21)
Recall (%)	0 (6)	75 (16)	70 (26)	**80 (30)**	**80 (22)**	74 (8)	79 (17)	69 (18)	98 (4)	66 (43)
Precision (%)	0 (61)	87 (2)	87 (5)	**87 (2)**	**89 (2)**	87.2 (0.7)	86 (4)	87 (2)	54 (6)	78 (19)

Table 3. Overview of the performance of the ten state-of-the-art classifiers on Test set (1) and (2). Median and interquartile range reported for the ten models obtained via ten-fold cross-validation.

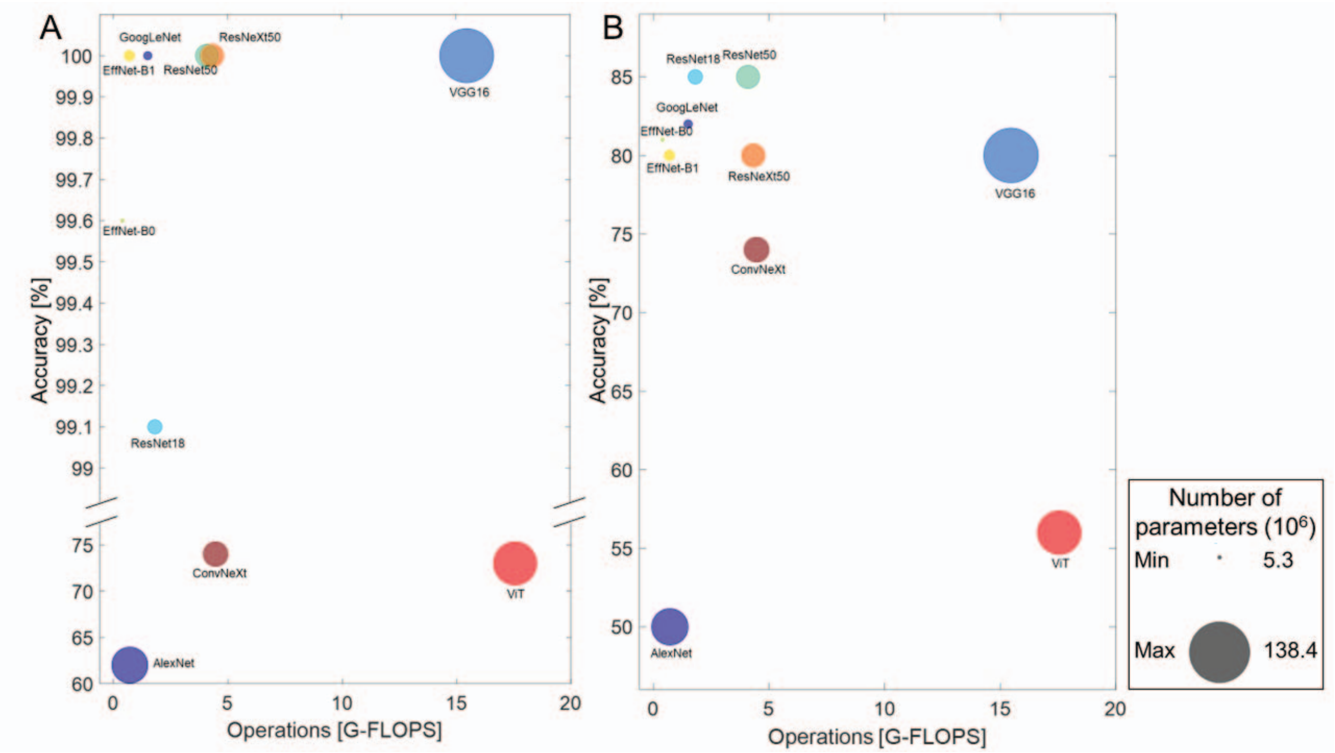

Figure 3. Ball chart reporting the median accuracy vs computational complexity. Accuracy using only the center crop versus floating-point operations (FLOPs) required for a single forward pass are reported. The size of each ball corresponds to the model complexity. A Test set (1); B Test set (2).

(1). This means that the N_2 disordered fruit could be identified. However, the detection of CO_2 disordered fruit was more difficult, as observed from the lower recall for test set (2).

5.4. Computational requirements

Besides the generalizability, an important aspect to use deep learning in a fast sorting system is the computational workload and speed. To this end, the accuracy on both test sets, the amount of G-FLOPS, and the number of parameters were evaluated. Table 3 and Figure 3 illustrate that GoogLeNet and ResNeXt provide excellent results on the first test set, while using a limited number of parameters. This means that the model efficiently uses its parameters. Other good architectures with high performance and limited parameters for this specific task are the ResNet and EfficientNet families. We consider the other models as inferior because of the lower accuracy (AlexNet, ViT, and ConvNeXt) or higher number of parameters (VGG 16). Looking at the robustness on test set (2), the ResNet family reached the highest performance, with ResNet50 slightly more demanding in computational power than ResNet18. EfficientNet and GoogleNet had even less computational requirements, but scored lower on performance.

To summarize, the results demonstrate that more parameters do not always lead to better performance. During model design, the architecture and data are critical in addition to the limited computational resources in an industrial setting. GoogLeNet, with an accuracy of 100 (0)% and 82 (2)% on the two test sets, a low number of parameters and FLOPs scored the best for our practical implementation of apple sorting based on X-ray radiographs. Also, ResNet50 and ResNeXt50 resulted in similar performances.

The X-ray system employed in this study provided data with a pixel size of 0.135 mm after binning. These images were cropped and resized finally resulting in a pixel size of 0.98 mm. This is an appropriate resolution as the disorders we want to detect are typically in the mm-range (Fig. 2). The speed of data acquisition can even be increased to 50 cm/s, which corresponds to 5-10 apples per second. X-ray imaging is proven to be safe on food for human consumption and is already used for foreign object detection. Such objects, such as metal, are easier to detect as the density differences between the food and the object is high. Small density differences inside fruit tissue due to disorders are more challenging and require advanced algorithms.

The study also came with some challenges. For instance, the Transformer-inspired models, i.e., ViT and ConvNeXt, were difficult to stabilize. Furthermore, the used training-validation dataset consisted of 1200 images which is very limited, especially for Transformer models [8]. The CNNs could yet reach a high performance on this small dataset. However, the detection of CO_2 disordered fruit should still be improved. In future research, the dataset should be further extended to enhance the model's performance. Apart from other types of internal disorders, the generalizability of the proposed deep learning model should also be checked for other seasons, cultivars, and data from an X-ray line scanner in another industrial set-up. In addition, more lightweight model architectures can be considered as well as explainable AI. For instance, heatmaps can be produced to better understand the model's attention into the images. Finally, the method can be extended to other species of fruit and vegetables.

6. Conclusions and Future Work

In this paper, we have presented a study of several deep learning architectures for classification on healthy and defect apples based on X-ray imaging data. The model that provides the best performance requiring limited computational resources is GoogLeNet. In future work, the training dataset could be increased to enhance model performance and robustness. Additionally, the method can be extended to other apple cultivars and species of fruit and vegetables. We conclude that the application of deep learning on X-ray radiography data for internal quality detection holds promising preliminary results for proper sorting in the agrifood industry. An industrial sorting system that implements the deep learning classifier will be able to quickly distinguish healthy and defect apples. This sample-by-sample quality control will result in decreased food losses compared to current batchwise inspection that cuts open a sample of fruit by hand.

7. Acknowledgement

The Research Foundation – Flanders (FWO grant Astrid Tempelaere 1S23721N, Hercules I013518N) and KU Leuven (project C3191059) are acknowledged for financial support.

References

[1] Mohammed Raju Ahmed, Jannat Yasmin, Eunsung Park, Geonwoo Kim, Moon S Kim, Collins Wakholi, Changyeun Mo, and Byoung-Kwan Cho. Classification of watermelon seeds using morphological patterns of x-ray imaging: A comparison of conventional machine learning and deep learning. *Sensors*, 20(23):6753, 2020. 1

[2] Etienne Belin, David Rousseau, Joël Léchappé, M Langlois-Meurinne, and Carolyne Dürr. Rate-distortion tradeoff to optimize high-throughput phenotyping systems. application to x-ray images of seeds. *Computers and electronics in agriculture*, 77(2):188–194, 2011. 1

[3] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* , pages 1251–1258, 2017. 1

[4] Clissia Barboza da Silva, Alysson Alexander Naves Silva, Geovanny Barroso, Pedro Takao Yamamoto, Valter Arthur, Claudio Fabiano Motta Toledo, and Thiago de Ara újo Mastrangelo. Convolutional neural networks using enhanced radiographs for real-time detection of sitophilus zeamais in maize grain. *Foods*, 10(4), 2021. 1

[5] Thomas De Schryver, Jelle Dhaene, Manuel Dierick, Matthieu N Boone, Eline Janssens, Jan Sijbers, Mattias van Dael, Pieter Verboven, Bart Nicolai, and Luc Van Hoorebeke. In-line ndt with x-ray ct combining sample rotation and translation. *NDT & E International*, 84:89–98, 2016. 2

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 2

[7] Gadgile Dhondiram and Chavan Ashok. Detection of postharvest fungal diseases of mango by x-ray scanning nondestructive technology. *Plant Pathology & Quarantine* , 7(1):65–69, 2017. 1

[8] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 1, 4, 5, 7

[9] EE Finney and KH Norris. X-ray scans for detecting hollow heart in potatoes. *American Potato Journal*, 55:95–105, 1978. 1

[10] Dhondiram Panditrao Gadgile, Chandrakant Padmakar Joshi, Vikas Madhukarrao Shinde, and Parshuram Babarao

Kachare. Detection of green mould rot infection of citrus fruit by x-ray scanning non-destructive technology. *Current Botany*, 8, 2017. 1

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 3, 5

[12] Els Herremans, Pieter Verboven, Evi Bongaers, Pascal Estrade, Bert E Verlinden, Martine Wevers, Maarten LATM Hertog, and Bart M Nicolai. Characterisation of 'braeburn'browning disorder by means of x-ray micro-ct. *Postharvest Biology and Technology*, 75:114–124, 2013. 1

[13] KC Kamal, Zhendong Yin, Mingyang Wu, and Zhilu Wu. Depthwise separable convolution architectures for plant disease classification. *Computers and electronics in agriculture*, 165:104948, 2019. 5

[14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 3

[15] Alex Krizhevsky. One weird trick for parallelizing convolutional neural networks. *arXiv preprint arXiv:1404.5997*, 2014. 3

[16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 1, 3, 5

[17] OL Lau. Effect of growing season, harvest maturity, waxing, low o2 and elevated co2 on flesh browning disorders inbraeburn'apples. *Postharvest Biology and Technology*, 14(2):131–141, 1998. 1

[18] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 1

[19] Li-Hua Li and Radius Tanone. Vision transformer approach for vegetables recognition. In *2022 International Seminar on Application for Technology of Information and Communication (iSemantic)*, pages 113–118. IEEE, 2022. 2

[20] Xiaopeng Li, Xiaoyu Chen, Jialin Yang, and Shuqin Li. Transformer helps identify kiwifruit diseases in complex natural environments. *Computers and Electronics in Agriculture*, 200:107258, 2022. 2

[21] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986, 2022. 1, 4, 5

[22] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 3

[23] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 3

[24] Ahmed M Rady, Daniel E Guyer, Irwin R Donis-González, William Kirk, and Nicholas James Watson. A comparison of different optical instruments and machine learning techniques to identify sprouting activity in potatoes during storage. *Journal of Food Measurement and Characterization*, 14:3565–3579, 2020. 1

[25] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018. 1

[26] Thomas F Schatzki, Ron P Haff, Richard Young, Ilkay Can, Lan Chau Le, and Natsuko Toyofuku. Defect detection in apples by means of x-ray imaging. *Transactions of the ASAE*, 40(5):1407–1415, 1997. 1

[27] Mahmoud A Shahin, Ernest W Tollner, and Ronald W McClendon. Ae—automation and emerging technologies: artificial intelligence classifiers for sorting apples based on watercore. *Journal of agricultural engineering research*, 79(3):265–274, 2001. 1

[28] Mahmoud A Shahin, Ernest W Tollner, Ronald W McClendon, and Hamid R Arabnia. Apple classification based on surface bruises using image processing and neural networks. *Transactions of the ASAE*, 45(5):1619, 2002. 1

[29] H Sherif, A Charrier, P Rasti, E Guiloteau, J Messaka, and D Rousseau. Automatic fasciation detection in salad with 2d x-ray imaging. In *XXXI International Horticultural Congress (IHC2022): III International Symposium on Mechanization, Precision Horticulture, and 1360*, pages 225–228, 2022. 1

[30] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1, 3, 5

[31] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015. 1, 3, 5

[32] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 1, 3, 5

[33] Astrid Tempelaere, Leen Van Doorselaer, Jiaqi He, Pieter Verboven, and Bart Nicolai. Fast sorting of defect apple fruit via x-ray imaging and artificial intelligence. VII International Symposium on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain, 2023. 4

[34] Astrid Tempelaere, Leen Van Doorselaer, Tim Van De Looverbosch, Michiel Pieters, Jiaqi He, Pieter Verboven, Ann Schenk, and Bart Nicolaï. Influence of different origins, ca storage conditions and storage times on internal browning in apple and pear cultivars. In *XXXI International Horticultural Congress (IHC2022): International Symposium on Postharvest Technologies to Reduce Food Losses 1364*, pages 213–220, 2022. 1

[35] Leon A Terry, Carlos Mena, Adrian Williams, Nigel Jenney, and Peter Whitehead. Fruit and vegetable resource maps: Mapping fruit and vegetable waste through the wholesale supply chain. 2011. 1

[36] Huy-Tan Thai, Kim-Hung Le, and Ngan Luu-Thuy Nguyen. Formerleaf: An efficient vision transformer for cassava leaf disease detection. *Computers and Electronics in Agriculture*, 204:107518, 2023. 2

[37] Mattias van Dael, Sekina Lebotsa, Els Herremans, Pieter Verboven, Jan Sijbers, UL Opara, PJ Cronje, and Bart M Nicolaï. A segmentation and classification algorithm for online detection of internal disorders in citrus using x-ray radiographs. *Postharvest Biology and Technology*, 112:205–214, 2016. 1

[38] Mattias van Dael, Pieter Verboven, Jelle Dhaene, Luc Van Hoorebeke, Jan Sijbers, and Bart Nicolai. Multisensor x-ray inspection of internal defects in horticultural products. *Postharvest Biology and Technology*, 128:33–43, 2017. 1

[39] Mattias Van Dael, Pieter Verboven, Angelo Zanella, Jan Sijbers, and Bart Nicolai. 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 Biology and Technology*, 148:218–227, 2019. 2

[40] Tim Van De Looverbosch, Jiaqi He, Astrid Tempelaere, Klaas Kelchtermans, Pieter Verboven, Tinne Tuytelaars, Jan Sijbers, and Bart Nicolai. Inline nondestructive internal disorder detection in pear fruit using explainable deep anomaly detection on x-ray images. *Computers and Electronics in Agriculture*, 197:106962, 2022. 1, 2

[41] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017. 1, 3

[42] Saikun Yu, Ning Wang, Xiangyan Ding, Zhengpan Qi, Ning Hu, Shuyong Duan, Zeqing Yang, and Xiaoyang Bi. Detection of pear freezing injury by non-destructive x-ray scanning technology. *Postharvest Biology and Technology*, 190:111950, 2022. 1