An automatic non-invasive classification for plant phenotyping by MRI images: An application for quality control on cauliflower at primary meristem stage

Yifan Zhou, Raphaël Maître, Mélanie Hupel, Gwenn Trotoux, Damien Penguilly, François Mariette, Lydia Bousset, Anne-Marie Chèvre, Nicolas Parisey

To cite this version:

Yifan Zhou, Raphaël Maître, Mélanie Hupel, Gwenn Trotoux, Damien Penguilly, et al.. An automatic non-invasive classification for plant phenotyping by MRI images: An application for quality control on cauliflower at primary meristem stage. Computers and Electronics in Agriculture, 2021, 187, pp.106303. 10.1016/j.compag.2021.106303 . hal-03347927

HAL Id: hal-03347927
https://hal.inrae.fr/hal-03347927v1

Submitted on 2 Aug 2023

Manuscript number COMPAG_2020_2352_R2

With: 2 Tables; 8 Figures (online only color); 1 Appendix

Title An automatic non-invasive classification for plant phenotyping by MRI images: an application for quality control on cauliflower at primary meristem stage

Yifan Zhou[a], Raphaël Maître[d], Mélanie Hupel[b], Gwenn Trotoux[a], Damien Penguilly[c], François Mariette[b], Lydia Bousset[a], Anne-Marie Chevre[a] and Nicolas Parisey[a],

[a]*Institute for Genetics Environment and Plant Protection (IGEPP), Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), Domaine de la Motte, 35653, Le Rheu, France*

[b]*INRAE OPAALE, 17 avenue de Cucillé, CS 64427- 35044 Rennes, France*

[c]*CATE - Station expérimentale, Vezendoquet, 29250 Saint Pol de Léon, France*

[d]*OBS - Organisation Bretonne de Sélection, Kernonen, 29250 Plougoulm, France*

Corresponding Author: Nicolas Parisey UMR1349IGEPP, INRAE Domaine de la motte 35 653 Le Rheu cedex, France Tél 02 23 48 51 95 Fax : 02 23 48 51 50 nicolas.parisey@inrae.fr

Highlights

* An automatic non-invasive method detects cauliflower curd deformation

* Tomographic images analysed by machine learning and deep learning methods

* Depending on the plant developmental stages, cross-validated F1-score were up to 95%

* On combined developmental stages, cross-validated F1-score is 88.67 %.

Abstract

During the past few years, milder autumn and winter seasons have caused severe problems to cauliflower harvest of Brittany region in France, mainly due to curd deformation. Consequently, cauliflower breeders are working on breeding new varieties that are more robust to climate change to stabilize the quality of cauliflower production. The aim of this study was to identify at which stage of the curd formation, significant difference can be detected between healthy and stressed cauliflower. A non-invasive classification based on Magnetic Resonance Imaging (MRI) images for cauliflower phenotyping was proposed. Plants exposed to vernalization stress were sampled at different times around primary meristem stage, then both MRI imaged and apex dissected. A work flow was developped to extract features from MRI images. A classification on phenotype was learned by LDA, QDA, PLSDA and CNN binary classification between two groups: healthy and stressed cauliflower. Promising F1 score and MCC up to 95% were achieved. Curd deformation is the main cause for cauliflower's later physiological disorders when reaching maturity. Therefore, the cauliflowers with deformation could be removed at the earliest, e.g., screening for plant breeding. At the same time, the healthy cauliflowers are not destroyed and continue their life cycle.

Keywords plant phenotyping ; non-invasive classification ; cauliflower primary meristem ; MRI application ; discriminant analysis

Abstract

During the past few years, milder autumn and winter seasons have caused severe problems to cauliflower harvest of Brittany region in France, mainly due to curd deformation. Consequently, cauliflower breeders are working on breeding new varieties which are more robust to climate change to stabilize the quality of cauliflower production. The aim of this study was to identify at which stage of the curd formation, significant difference can be detected between healthy and stressed cauliflower. A non-invasive classification based on Magnetic Resonance Imaging (MRI) images for cauliflower phenotyping was proposed. Plants exposed to vernalization stress were sampled at different times around primary meristem stage, then both MRI imaged and apex dissected. A work flow was developped to extract features from MRI images. A classification on phenotype was learned by LDA, QDA,PLSDA.

Statistical analysis was then applied for a binary classification between two groups: healthy and stressed cauliflower. Promising F1 score and MCC up to 95% were achieved. Curd deformation is the main cause for cauliflower's later physiological disorders when reaching maturity. Therefore, the cauliflowers with deformation could be removed at the earliest, e.g., screening for plant breeding. At the same time, the healthy cauliflowers are not destroyed and continue their life cycle.

Keywords: plant phenotyping ; non-invasive classification ; cauliflower primary meristem ; MRI application ; discriminant analysis

1 Introduction

According to the Food and Agriculture Organization of the United Nations, large-scale experiments in crop phenotyping are a key factor in meeting the future agricultural needs to feed the world and provide biomass for energy while using less water and fertilizer under a constantly evolving environment adapted to climate change (Minervini et al., 2015). However, current assessments of phenotypic characteristics for disease resistance or stress in breeding programs rely largely on visual scoring by experts, which is laborious and dull, not sufficiently objective or destructive (Busemeyer et al., 2013). Various imaging methodologies are being used to collect data for quantitative studies of complex traits related to growth, yield and adaption to biotic or abiotic stress (disease, insects, drought and salinity) (Li et al., 2014). There is an urgent need to develop reliable computer vision methods that can extract phenotypic information from experiments at scales from single cell to whole plant, in the greenhouse or on the field (Li et al., 2014). The extracted information, integrated with genetic and environmental data by novel models based on accurate, robust and automatic statistical analysis will give new opportunity to genetic diversity screening, new breeding strategies in agriculture as well as market management.

Given the rapid development of high-throughout genotype screening in plant breeding and genomics for related growth, yield and tolerance to different biotic and abiotic stresses, there is a call for more effective and reliable phenotyping data to support modern genetic crop improvement (Li et al., 2014). To accomplish this goal, more and more projects for plant phenotype unite expertise from biological science, computer science, mathematics and engineering. Such an approach was needed to offer cauliflower breeders efficient screening methods on plant development as early as possible, by associating plant phenotype with genomes in imaging systems of computer vision.

In recent years, cauliflower winter harvest for the region of Brittany in France has been observed to be very unstable and extremely reduced due to warmer autumns (Tremellat, 2017). Physiological disorders, such as open, ricey (Watts, 1966) or bracty (Kop et al., 2003) head appear. A healthy head is tightly compact with only florets and forms one bracts (Fig. 1a), whereas an open head has gaps among florets (Fig. 1b); a ricey head has protruding flower buds (Fig. 1c); and a bracty head has leaves intermingled with florets (Fig. 1d). These deformations renders cauliflower heads unmarketable, resulting in important commercial losses, e.g. with about a third of harvest was unmarketable (Tremellat, 2017).

To induce flowering and thus curd formation, beyond the juvenile phase, cauliflower must be exposed to vernalization at "relative cold" temperature (Wurrand Fellows, 2000). In the couple of weeks following vernalization, cauliflower primary meristem undergoes curd formation period, divided into 4 stages (Kieffer et al., 1998): vegetative, curd-induction, curd-forming and curd-thickening stage, noted as Stage 1, 2, 3 and 4, respectively in this paper. This curd formation period is critical for cauliflower growth. If deformation appears during this period, the cauliflower head will remained deformed during subsequent growth until maturity, about 2 months later. Breeding cauliflowers less sensible to autumn temperature fluctuation is thus desirable to stabilize yields in autumn to winter harvests. To render such breeding possible, early stage phenotyping during curd formation is needed.

Floral initiation in cauliflower is the result of fine regulation of a whole network of genes and regulatory loops with interplay between transcription factors (Goslin et al., 2017). This regulation interacts with vernalization (Matschegewski et al., 2015). Bracting in cauliflower depends on its genotype (Kop et al., 2003) and the climate during floral initiation. The effect of temperature and developmental stage on bracting and riciness quality defects have already been studied in the field, either during harvest time (Grevsen et al., 2003) or by a destructive sampling with scanning electron microscopy (Fujime and Okuda,1996).

Using non-invasive methods, healthy cauliflower without deformation could be kept for further growth. However, at this moment, the apex of cauliflower meristem is only around 0.5mm of diameter, still tightly wrapped in a bunch of huge leaves making it invisible to naked eyes (Fig. 2) and preventing the use common RGB cameras to capture high resolution pictures on meristem without destroying the plant. For organs inavailable from the outside as plant apex, among the available techniques one could either use external imaging e.g. spectrometry or hyperspectral imaging; or internal imaging e.g. Xray or MRI. External imaging would be possible if the external part of the plant has features correlated to the apex deformation. At such early stages as the one investigated in this study, experts are not able to assess the healthy state by external observation of plant morphology. Raman spectrometry has been investigated in a companion project but will not be discussed here as no success was achieved. Current internal imaging techniques are Xray and Magnetic Resonance Imaging (MRI). Xray relies on a difference in tissue density, whereas MRI is based on the relaxation time of the tissues. Specifically, in plants the structure and chemical composition of the tissues imfluences the water molecules relaxation time (Musse and Van-

As, 2018). Even though Xrays are cheaper and easier to use, this makes MRI a more selective technique to detect differences in tissues structure among plants. Although in most applications MRI is used to investigate water relations and transport in plant tissues, MRI can also be used to measure other plant constituents, such as metabolites and air spaces. MRI therefore provides access to a wide range of information about plant tissues, including structural characteristics at different length scales and physical-chemical features (Musse and Van-As, 2018). We hypothetised that such changes might occur in the transition from vegetative to floral induction in plant apex.Recent advances in spectroscopic techniques, such as Magnetic Resonance Imaging (MRI) promise non-invasive measurements on plant (Rascher et al., 2013) allowing assessment of plant phenotype revealing plant's inner part without destroying its outside part. However, due to apex's tiny size, its MRI image has only few pixels, which prevents an efficient image analysis. As the stems with deformed meristem have a form different from that of healthy one (Hupel, 2018), this offers the prospect of phenotyping by image analysis of stem shapes.

The aim of this study was to identify at which stage of the curd formation, significant difference can be detected between healthy and stressed cauliflower. Plants exposed to vernalization stress were sampled at different times around primary meristem stage, then both MRI imaged and apex dissected. A work flow was developped to extract features from MRI images. Statistical analysis was then applied for a binary classification between two groups: healthy and stressed cauliflower. Since this application is designed to satisfy industrial need, focused was set on how to solve practical problems encountered during different steps in work flow, at same time, proposing adequate and efficient models.

2 Materials and methods

2.1 Data collection

Two environment conditions were imposed in greenhouse to simulate a normal autumn (group H) and a stressful warmer autumn weather (group S). The latter one was to simulate a milder autumn leading to a shorter vernalization period for cauliflower. A F1 hybrid variety of cauliflower detected as sensitive to temperature fluctuations when cultivated at large-scale Brittany region of France was selected for the experiment. Cauliflower seeds were sown in June 2018 on a mixture of sand and vermiculite at 20°C, with plants in group S sown two weeks later than plants in group H. Their seedlings at two-leaf stage were transferred to a plastic greenhouse at a temperature of 12-13°C for plant hardening during 7 weeks. The plants were transplanted in 7.5L pots and and grown for 7 further weeks, then vernalized at 4°C for either 2 weeks (group S) or 4 weeks (group H).

After the vernalization, plants in group H were cultivated under tunnels with an average daily temperature between 9 and 15°C preventing the risk of devernalization, whereas those in group S were kept in a greenhouse with a temperature between 15 and 20°C simulating a warmer autumn. In this way, it was expected that the seeds of the same homogeneous variety would grow to healthy cauliflower heads under normal autumn, called group H, but those under stressful autumn would grow to stressed ones, called group S, in this paper.

The plants were sent sampled for MRI measurement weekly between 0 and 31 days after the end of vernalization, 5 plants per date and per treatment each time, in order to have samples distributed on the 4 stages of primary meristem.

The MRI measurement was carried out by a 1.5 Tesla MRI whole body scanner (Magnetom Avanto, Siemens, Erlangen, Germany) equipped with an eight-channel "knee" receiver coil. The plant was laid down on the examination table because it was too high to be placed uprightly (Fig. 3). Due to this reason, cauliflower apex might not be found in the coil center due to gravity. This was especially true for plant of meristem at stage 1 and 2 due to its young and fragile stem. Another practical issue was that plant leaves were piled up in the ring resulting in aliasing artifacts in MRI. Therefore, some extra image pre-processing methods were carried out and will be explained in Section 2.2.1.

The 3D MRI images were acquired by a 3D turbo spin echo (TSE3D) sequence with a voxel size of $0.5mm\times0.5mm\times0.5mm$, a matrix size of 192×192, a FOV of 96×96, 96 slices per volume, a slice thickness of 0.5mm, an echo time of 9.5ms, a repetition time of 500ms, 2 averages, a turbo factor of 14 and a bandwidth of 263 Hz/pixel. These values were chosen as the best compromise between on the one hand enhancing, at best, the contrast between the stem and the rest of the plant by the MRI operator at the time of acquisition and on the other hand, the acquisition time short enough to allow analysis of a sufficient number of plants per day.

The acquisition time for one plant was about 33 min. Plant were well watered at the eve of every acquisition in order to improve MRI image contrast.

After the MRI acquisition, cauliflower was dissected and photographed to enable breeding experts to identify growth stage and thus construct a ground truth for the database. Dissected cauliflower apex were stained with aceto-carmine, observed under a magnifying lens (Nikon SMZ-U, zoom 1:10) and digital RGB images were taken (Fig. 4). In this article, four stages were distinguished: 1. Vegetative stage; 2. Curd-induction stage; 3. curd-forming stage; 4. Curd-thickening stage.The top line (images ABCD) in Figure 4 illustrates a schematic representation of cauliflower floral induction based on the scanning electron micrographs in (Kieffer et al., 1998). At vegetative stage, only leaf scales are produced (Fig. 4A). During the curd-induction stage, an enlarging empty area becomes visible between leaf scales (Fig. 4B). Through the curd-forming stage, round floral primordia appears at axil of each bract scale (Fig. 4C). The growth of bracts scales is repressed at curd-thickening stage (Fig. 4D). For further growth, the apex consists of only floral primordia stopped in their development (Smyth1995), grown into florets (ramified group of flowers). Their further maturation into flowering is postponed long afterwards.

The middle (images EFGH) and bottom (images IJKL) line in Figure 4 give corresponding examples of each of the stages on healthy and stressed cauliflower, respectively. Altogether, 100 plants were collected for the year 2018 and 60 plants for 2019.

These RGB images were used for a double-blinded identification of meristem developmental stage, compared manually with Fig. 4 by experts on cauliflower's mersitem morphological development in order to decide their corresponding sample's primary meristem stage.

2.2 Feature extraction

For each dissected plant, called a sample in this paper, a set of MRI images was acquired, called raw images. Feature extraction consisted of three steps: extraction of region of interest, image of contour and image of skeleton.

2.2.1 Extraction of region of interest

Its aim was to select slice from raw images (Fig. 5) and pre-process the selected slice. There were 96 raw images on plane XY per sample. The raw image resolution was 192×192 pixels. Therefore, a sample can be represented in 3D (Fig. 5b) with the apex of cauliflower meristem circled in red. In this 3D presentation, the top of a cauliflower plant has of volume of 0.5dm^3. Reminding that an apex was only about 0.5mm, it barely corresponded to one pixel on raw image. Hence, it was chosen to extract features on cauliflower's stem appearance.

Because variable numbers of leaves had been included in the coil during MRI acquisition, rather than extracting 3D morphological features directly from raw images, the plant 3D morphological information was conrained withing the 2D images on plane XZ (Fig5c) and YZ around the stem apex (Fig5d). Hence, the database consisted of 320 slices for 160 samples.

If the plant apex was perfectly in coil center during MRI acquisition, the two middle slices on plane XZ and YZ were the best illustration. However, when the plant apex was not perfectly in coil center, the cauliflower meristem was missed or occluded in middle slice (Fig. 6a,b) Therefore, a manual selection was required in order to find one best slice on each plane which illustrated the meristem as clearly and entirely as possible (Figure 6c,d).

Aliasing artifacts of MRI acquisition happened when one leave exceeding scanner's field of view was partially projected onto the other side of image, (Fig. 6e). Aliasing is an artifact that can occur in MRI images acquisition when the scanned object is larger than the square image area that is to be measured, which is called the field of view (FOV). As a consequence of sampling issues, portions of the object outside of the desired FOV get mapped to an incorrect location inside the FOV. For our 2D slices, this artefact reproject leaves with much higher contrast than the rest of the stem. In order to work properly on the stem, beside the presence of this artefact, we improve the local contrast and enhance the definitions of edges everywhere in the slice by using histogram equalization. Hence, a contrast limited adaptative histogram equalization (CLAHE) that enhance contrast by using information in the vicinity of each pixel while puting limitations (i.e. maxima) to the extent of said constrast augmentation (Gonzalez and Woods, 2008) was applied on this kind of slices (Fig. 6f).

To focus on cauliflower stem apex morphology, and remove unnecessary plant leaves and petioles, an extraction of region of interest (ROI) was performed as follows Fig. 7a to 7c). A contour extraction method based on Otsu's thresholding (Otsu, 1979; Gonzalez and Woods, 2006) was firstly used to find object of interest, generating an image called mask in this paper (Fig. 7b). The slice on the mask was scaled to a higher resolution, from 192×96 to 384×192 pixels. To improve contrast inside ROI, the Contrast Limited Adaptive Histogram Equalization algorithm (CLAHE) (Pizer et al., 1987) was carried out on scaled image. This final ROI image (Fig. 7c), was ready for further feature extraction.

2.2.2 Image of contour

To extract morphological features on cauliflower stem apex, the same contour method (Gonzalez and Woods, 2006) was again applied on the ROI image. Only the contour with the largest area was considered as the final object of interest (Fig. 7d), called image of contour in this paper. Five categories of features were calculated: contour marked in green, rectangle in blue, hull, ellipse in red and intensity (Fig. 7d). For each category, 3 or 5 features were computed (Table 1), chosen in a way that the features' value was invariant to object position in the image.

2.2.3 Image of skeleton

To extract morphological features on cauliflower main stem, an image of skeleton was produced as follows. From a ROI image (Fig. 8a), the corresponding mask (Fig. 8b) was first morphologically thinned with a maximum iteration of 10 by (Zhang and Suen, 1984), (Fig. 8c). It was then skeletonized to 1 pixel (Fig. 8d). The difference between the thinned and skeletonized image was the image of skeleton (Fig. 8e), having a similar form to plant main stem. Three categories of features were calculated on image of skeleton: contour marked in green, rectangle in blue and hull (Table 1).

2.3 Binary classification with discriminant analysis

The classification issue was binary, with the two groups of cauliflower either healthy or stressed. The database consisted of 160 plants, with 320 images, distributed on 4 stages (Table 2, see columns Stage, Size, H, S) with nearly half of plants in stage 4,. Due to this limited data size, a **leave-one-out cross-validation** (Devijver and Kittler, 1982) was applied to assess the predictive capability of the classifiers.

Several supervised algorithms were tested for learning and validation steps based on features extracted from image of contour or/and image of skeleton: Linear Discriminant Analysis (LDA) (Fisher, 1936), Quadratic Discriminant Analysis (QDA) (Hastie et al., 2009) and Partial Least Squares Discriminant Analysis (PLSDA) (Barker and Rayens, 2003). Simple classifiers were chosen because of covariates multicollinearity, small samples and unbalance in our dataset can be troublesome to highly non-linear and/or complex classifiers.

The application had significant ratios of feature number to sample size. For example, the number of features for the chosen classifiers on image of contour and skeleton on stage 3 is 34 versus 26 slices from 13 samples. In order to avoid overfitting and multicollinearity problems in marchine learning (Burnham and Anderson, 2002), LDA was used with automatic shrinkage by Ledoit-Wolf lemma (Ledoit and Wolf, 2004), QDA with regularized covariance (Friedman,1989) and we used PLSDA.

The regularization parameter for QDA and the adaptive component number for PLSDA were automatically chosen by a nested leave-one-out cross-validation with the inner layer to find hyper-parameters giving best F1 score on subsamples and the outer layer to evaluate algorithm performance on the whole cross-validated data.

F1 score (F1) (Rijsbergen, 1979) Matthews Correlation Coefficient (MCC) (Matthews, 1975) and Jaccard Index (JI) (Jaccard, 1912) were used to evaluate the classifiers performance. The MCC was a balanced measure of the quality of binary classification with +1 a perfect prediction, 0 no better than random prediction and -1 a total disagreement between prediction and observation. The JI was defined as the size of the intersection divided by the size of the union of two label sets which are here the predicted set of labels and the observed set of labels. These metrics are adapted to the situation where two classes were of very different size (Chicco and Jurman, 2020) (Table 2, see different size between "H" and "S")).

2,4 Binary classifications with deep learning

To complement our classifications based on selected features, deep learning was implemented, a class of computational models composed of multiple processing layers learning representations of data with multiple levels of abstraction (LeCun et al., 2015). Those algorithms have proved to be very efficient in a wide variety of domains, most notably computer vision (Emmert-Streib et al., 2020). Among those algorithms, Convolutional Neural Networks (CNNs) are well known for their success in many computer vision tasks such as image classification (Krizhevsky et al., 2012) and objects recognition (Li et al., 2015). Deep learning was used to explore two main questions. First, could a CNN, using slices, lead to higher scoring than our classifications based on selected features on said slices? Second, could a CNN, using directly the 3D volumes, lead to good scoring?

To explore the ability of a CNN to outperform other classifiers, on selected slices, transfer learning was used, a technique to re-purpose a previously trained model (Yosinski et al., 2014). A classical Xception architecture (Chollet et al, 2017) was used as base model, pretrained on the large generalist ImageNet dataset (Deng et al., 2009) with more than 14 million images of thousands of categories. The fully connected layers were removed (and associated multiclass problem) and the rest of the convolutional layers were used as fixed feature extractors to feed a new neural network of two fully connected layers of 512 neurons. From here on, this CNN will be referred to as the 2D CNN. After the first transfer-learning step, subsequent fine-tuning i.e. training also the convolutional layers was also tried. In both cases, given our small datasets, on-the-fly data augmentation with rotations and axial symmetries was used.

To assess the properties of a CNN using 3D volumes, a readily available architecture previously used on CT Scans for binary classification in human epidemiology (Zunair et al., 2020) was adapted. From here on, this CNN will be referred to as the 3D CNN. Due to the relative scarcity of volumic datasets and their divergence from our use case, transfer learning was not chosen and the 3D CNN was trained from scratch. Again, on-the-fly data augmentation with rotations was used. All deep learnings were performed by dividing our dataset into training (70%) and validation (30%) sets.

2.5 Hardware and librairies used

The codes in the application were written in python using library OpenCV (Bradski, 2014), skimage (van der Walt et al., 2014), scikit-learn (Pedregosa et al. 2011) and tensorflow (Abadi et al., 2015). The calculation was carried out on a common desktop Dell Precision Tower 3420 with Intel Xeon E3-1225 v6, 8192KB cache and an NVIDIA Tesla K80.

The processing time for feature extraction was quite negligible, no more than several millisecond per image. The computation time for supervised learning and validation depended on sample and feature size. For example, it took less than 1 second for a cross-validation by LDA with features of image contour and skeleton (34 features) on stage 234 (232 images for 116 samples). Computation time for CNNs were up to an hour for 3D CNN over 200 epochs. For all our application, from pre-processing steps to deep learning, codes and a sample dataset are available upon request to the corresponding author.

3 Results

In order to decide from which stage the classifiers can distinguish healthy from stressed cauliflower apex, the classifiers were firstly learned and validated on samples of the year 2018 and 2019 together, but on separate individual stages, 1, 2, 3 or 4 (Table 2 upper lines). On all the individual stages except on stage 1 (when cauliflower meristem was still on vegetative state), one or several classifiers could reach expectation (Boughorbel et al., 2017) with F1 above 85% and MCC above 65% (Table 2, marked in italics; see below for the rationale behind our expectation threshold). Therefore, the classifiers could distinguish healthy from stressed cauliflower as early as from curd-induction stage.

One or 2 classifiers with best performance on every stage are marked with an asterisk. LDA gives most of the best performance (marked with an asterisk) compared to QDA and PLSDA (Table 2). Besides, most of the best performance by LDA are computed by features from image of contour (Table 2). As the computation of the JI do not lead to significantly different conclusions than the ones based on F1 and MCC, in order to keep Table 2 comprehensible, we made JI scores available in Table 1 of the Appendix.

All the results reaching expectation are marked in italics, with F1 above 85% and MCC above 65% (Boughorbel et al., 2017). In the referenced article (Boughorbel et al, 2017), there are classifiers built on many different data sets (Table 5; 46 data sets). For a generalist classifier (in their case, the « SVM.imb » column), the average MCC obtained is 67.76%. We rounded roughly to 65% and considered that a good result would be strictly above this approximated mean score. For the F1-score, under the assumption of a class-balanced dataset, the score for a naïve classifier, that always predict the same class whatever the covariate values, is 66 %. If we consider such score as the least we can do and given the different levels of inbalance in our different data sets, which affect such minimal F1-score, we took a higher threshold, namely at 85%.

Based on the previous observation, classifiers on mixed stage 234 were computed in order to test whether features were sufficiently different to predict cauliflower healthy state without prior knowledge on meristem stage. The result was quite promising, with 88.67% on F1 and 67.93% on MCC based on features of contour and skeleton by QDA (Table 2 middle lines).

Since the classifiers were calculated by supervised algorithms in machine learning, we wanted to know the influence of sample size on classifier performance, by comparing the largest data set to smaller ones. For this account, classifiers were computed only on samples of the year 2018 on stage 4 and 234 (Table 2, botom lines). In this comparison, the sample size is increased by about 12% and 35% compared to samples of year 2018 and 2019, respectively. With more samples, the classifier performance was improved, between 0.3% to 10.3%. Therefore, it was concluded that the sample size did have an impact on the classifiers' performance.

Concerning deep learning, both the 2D CNN and the 3D CNN were trained to solve our binary classification problem on stage '234' from 2018 and 2019. This was our biggest relevant dataset even if it was still extremely small by deep learning standards.

For the transfer learning of the 2D CNN, a small learning rate (1e-3) was chosen and had to stop after a few epochs (20) before overfitting appeared as assessed by visual diagnostic of the loss functions between trained and validated datasets. For the subsequent fine-tuning of the 2D CNN, an even smaller learning rate (1e-5) was used and more epochs (50).

With transfer learning, the performance of the 2D CNN on the validation dataset were an F1-score of 73.19 % and a MCC of 46.39 %. With subsequent fine-tuning, there were small performance improvement with an F1-score of 75.48 % and a MCC of 50.97 %.

For the training of the 3D CNN, a low learning rate of 1e-5 was used for 50 epochs, afterward overfitting appeared. The performances of the 3D CNN on the validation set were an F1-score of 81.46 % and a MCC of 62.38 %.

4 Discussion

In this paper, a non-invasive classification for cauliflower phenotyping by MRI images was proposed. It was an application of screening on cauliflower still at primary stage long before its physiological disorders become visible to naked eyes. More specifically, MRI images were firstly acquired on cauliflower plant with its apex diameter of about 0.5mm, on which features by contour or/and skeleton were extracted. These features were then sent for learning by several discriminant analysis, such as LDA, QDA and PLSDA. The healthy state of cauliflower meristem influenced by temperature fluctuation was then predicted. If deformation had already occurred and would be sufficiently marked, dissecting the apex could give access to deformation. However, it could be the case that the molecular processes have already occurred but the deformation is not yet visible on the apex. Contrasting the two sets of temperature enabled us to detect morphological differences prior to the meristem deformation. However, within each of the two temperature groups, there was variation in the development, making it necessary to assess the actual growth stage of each individual plant. As an alternative to discriminant analysis on selected features, deep learning methods were used to generate predictions.

The classifiers could distinguish healthy or stressed cauliflower as early as from curd-induction stage. Experiments showed that cauliflower meristem developed very quickly into stage 1, just in few days after the end of the vernalization period. It might even occur during vernalization if the temperature was not cold enough. Therefore in practice, there would not have many cauliflower on stage 1 for the application on screening, usually scheduled several days after vernalization. Hence, the poor performance on stage 1 of the classifiers was negligible in an industrial context.

Classifiers on mixed stage 234 were computed in order to test whether features were sufficiently different to predict cauliflower healthy state without prior knowledge on meristem stage. The result was quite promising, with 88.67% on F1 and 67.93% on MCC based on features of contour and skeleton by QDA. Mixing curd induction, forming and thickening stage together could make the application even more automatic, since the only need was to decide a starting day for screening by avoiding meristem vegetative period, e.g, 5 days after the end of vernalization. In this manner, even though cauliflower grows differently due to temperature fluctuation, whether its meristem on stage 2, 3 or 4, will not have important impact on the performance of the prediction. Deep learning results were also promising on mixed stage. Although they were not able to outperform classifiers trained on selected features, their performances were very close. Another interesting result is that the deep learning using volumic data has the best performance among the architecture we tried. It showed on our (unusual) use case the versatility of these approaches.

Classifiers which were not only efficient but also with stable performance were needed. LDA gave most of the best performance (Table 2, marked with an asterisk) compared to QDA and PLSDA. In several of the experiments reported in Table 2, the F1 or MCC score was lower for the classification using image of contour and skeleton as compared to only image of contour. It could be surprising that the performance of the classifier decreases when more information becomes available to learn from. However, all of our classifiers are using mathematical regularizations (in a way or another) which leads, notably, to some features being nullified (their coefficients set to, nearly, zero) if they do not bring new information to solve the problem. This explains, in part, why the number of available features does not always lead to a high improvement in classification. Besides, most of the best performance by LDA are computed by features from image of contour. Therefore, applying classifier using features from image of contour by LDA for the application of screening on cauliflower at primary meristem stage is suggested.

Classifier performance was improved with more samples, between 0.3% to 10.3% (Table 2). Therefore, it was concluded that the sample size did have an impact on the classifiers' performance. However, it is likely that discriminant classifiers performances would reach a limit lower than that of state-of the art deep learning performances if enough data were available. Given that data scarcity is a bottleneck, a way forward would be to develop transfer learning and/or fine-tuning of a 3D CNN in a future work.

The experimental results showed that models had a rather promising performance on F1 score and MCC, especially for the one on features of contour by LDA. The classifiers could provide breeders with elements to decide whether to remove those stressed plants before planting. The associated environmental cost from negative by-product of cultivation on unmarketable plants, such as pesticide, soil resource, water and other energy could be saved.

The parameters for the classifiers were calculated completely in an automatic way. The only human intervention was during the extraction of region of interest step, that was to manually re-select slices from MRI raw images when plant did not grow straight upside and to add an histogram equalization on slices when MRI aliasing artefacts occurred. In practice, this manual selection largely depended on whether cauliflower's main stem was strong enough to support plant weight against gravity when laying down. It might be related to meristem young stage, like vegetative and curd-induction stage, but not absolutely. For example, 45% slices were re-selected for data in 2018 versus 98% for 2019. Even with the same genetic type, the plant might grow differently between years due to variable environmental conditions. Given the experiment with 3D CNN, if its performances were to be improved, it would be a good solution to remove the few remaining human interventions.

This manual selection problem might also be solved by constructing portable MRI system because image acquisition system's position can be adjusted to plant instead of laying down plant to adapt to fixed system. In fact, MRI bulky, costly and complex hardware limitation is the main factor which prevents it becoming a standard research tool in plant phenotype regardless of its non-invasive advantage. Nevertheless, several MRI mobile prototypes have been constructed from laboratories for potential industrial applications, such as measuring dynamic water change in living stems or fruit (Windt and Blümler, 2015). A relative low-cost wide bore MRI scanners have also been designed and constructed for rapid quality inspections of fruits and vegetables in order for an industrial food quality assurance and control (McCarthy and Zhang, 2012) (Milczarek and McCarthy, 2012). Hence, constructing mobile MRI systems is a potential research direction of future work.

Depending on the plant developmental stages, cross-validated F1-score were up to 95% and on combined developmental stages, cross-validated F1-score was 88.67 % (Table 2) and 81,46 % for deep learning. Yet only on one sensor (the MRI) was used for this study, another direction for improvement would be combining multi-modal acquisitions from different sensors, such as chlorophyll fluorescence (Rousseau et al., 2013), with ensemble methods (Zhou, 2012).

5 Conclusion

We aimed at improving the early detection of cauliflower curd deformation, the main cause for cauliflower's later physiological disorders when reaching maturity. A non-invasive classification based on Magnetic Resonance Imaging (MRI) images for cauliflower phenotyping was proposed, with tomographic images analysed by machine learning and deep learning methods. Promising F1 score and MCC up to 95% were achieved. Therefore, the cauliflowers with deformation could be removed at the earliest, e.g., screening for plant breeding. At the same time, the healthy cauliflowers are not destroyed and continue their life cycle. We consider this work as another proof of the usefulness and potential of tomographic data for non-invasive plant phenotyping

6 Acknowledgements

This work was part of the project ’GWASSICA’ co-funded by the Région Bretagne (France) and EU funds from the FEDER Program (Grant EU00430). A special thank to Mona Lostanlen from OBS Innovation for her meticulous work in cultivation, transportation and dissection. We thank Pascale Le-Neve and Patricia Nadan from INRAE Le Rheu for their administrative help.

7 CRediT author statement

Conceptualization: AMC, NP, LB, FM; Funding acquisition: AMC, NP, FM; Project administration: NP, AMC, RM, MH; Supervision & Validation: NP, AMC, LB; Investigation: RM, MH; Data curation: YZ, RM, MH, GT, DP; Formal analysis, Methodology: YZ, NP; Software: YZ, NP; Visualization, Writing - original draft: YZ; Resources, Writing - review & editing: all authors.

8 Declaration of Interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y. Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. arXiv:1603.04467.

Barker, M., Rayens, W., 2003. Partial least squares for discrimination. Journal of Chemometrics 17, 166 – 173.

Boughorbel, S., Jarray, F., El-Anbari, M., 2017. Optimal classifier for imbalanced data using matthews correlation coefficient metric. PLOS ONE 12, 1–17.

Bradski, G., 2000. The OpenCV Library. Dr. Dobb's Journal of Software Tools .

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2 ed., Springer-Verlag, New York.

Busemeyer, L., Mentrup, D., Möller, K., Wunder, E., Alheit, K., Hahn, V., Maurer, H.P., Reif, J.C., Würschum, T., Müller, J., Rahe, F., Ruckelshausen, A., 2013. Breedvision - a multi-sensor platform for non-destructive field-based phenotyping in plant breeding. Sensors 13, 2830–2847.

Chicco, D., Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21, 6.

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In : Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 248–255

Devijver, P., Kittler, J., 1982. Pattern Recognition: A Statistical Approach. Prentice-Hall, Englewood Cliffs.

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., Dehmer, M., 2020. An introductory review of deep learning for prediction models with big data. Frontiers in Artificial Intelligence 3.

Fisher, R., 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics 7, 179–188.

Friedman, J.H., 1989. Regularized Discriminant Analysis. Journal of the American Statistical Association 84, 165.

Fujime, Y., Okuda, N., 1996. The physiology of flowering in brassicas, especially about cauliflower and broccoli, 247–254.

Gonzalez, R.C., Woods, R.E., 2006. Digital Image Processing (3rd Edition). Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

González, R.C., Woods, R.E., 2008. Digital image processing, 3rd Edition. Pearson Education.

Goslin, K., Zheng, B., Serrano-Mislata, A., Rae, L., Ryan, P.T., Kwaniewska, K., Thomson, B., Ó'Maoiléidigh, D.S., Madueño, F.,Wellmer, F., Graciet, E., 2017. Transcription factor interplay between leafy and apetala1/cauliflower during floral initiation. Plant Physiology 174, 1097–1109.

Grevsen, K., Olesen, J., Veierskov, B., 2003. The effects of temperature and plant developmental stage on the occurrence of the curd quality defects "bracting" and "riciness" in cauliflower. Journal of Horticultural Science and Biotechnology 78, 638–646.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of Statistical Learning. Springer-Verlag New York, NY, USA.

Hupel, M., Décembre, 2018. Projet GWASSICA - Suivi de l'initiation florale et de la croissance de la pomme de chou-fleur par IRM. Technical Report. IRSTEA Rennes.

Jaccard, P., 1912. The Distribution of the flora in the Alpine zone.1. New Phytologist 11, 37–50.

Kieffer, M., Fuller, M., J. Jellings, A., 1998. Explaining curd and spear geometry in broccoli, cauliflower and 'romanesco': Quantitative variation in activity of primary meristems. Planta 206, 34–43.

Kop, E., Teakle, G., McClenaghan, E., Lynn, J., King, G., 2003. Genetic analysis of the bracting trait in cauliflower and broccoli. Plant Science 164, 803–808.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Ledoit, O., Wolf, M., 2004. Honey, I Shrunk the Sample Covariance Matrix. The Journal of Portfolio Management 30, 110–119.

Li, H., Lin, Z., Shen, X., Brandt, J., Hua, G., 2015. A convolutional neural network cascade for face detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5325–5334

Li, L., Zhang, Q., Huang, D., 2014. A review of imaging techniques for plant phenotyping. Sensors (Basel) 14, 20078–20111.

Matschegewski, C., Zetzsche, H., Hasan, Y., Leibeguth, L., Briggs, W., Ordon, F., Uptmoor, R., 2015. Genetic variation of temperature regulated curd induction in cauliflower: elucidation of floral transition by genome-wide association mapping and gene expression analysis. Frontiers in Plant Science 6, 720.

Matthews, B., 1975. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. Biochimica et Biophysica Acta (BBA) - Protein Structure 405, 442 – 451.

McCarthy, M., Zhang, L., 2012. Food Quality Assurance and Control.

Milczarek, R., McCarthy, M., 2012. Cheminform abstract: Low-field mr sensors for fruit inspection. ChemInform 43.

Minervini, M., Scharr, H., Tsaftaris, S.A., 2015. Image analysis: The new bottleneck in plant phenotyping [applications corner]. IEEE Signal

Musse, M., Van As, H., 2018. NMR imaging of air spaces and metabolites in fruit and vegetable. Modern Magnetic Resonance, Webb, G.A., Springer International Publishing, Cham, CHE, p. 1765-1779

Otsu, N., 1979. A threshold selection method from gray-level histograms.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K., 1987. Adaptive histogram equalization and its variations. Computer Vision, Graphics, and Image Processing 39, 355 – 368.

Processing Magazine 32, 126–131.

Rascher, U., Blossfeld, S., Fiorani, F., Jahnke, S., Jansen, M., Kuhn, A., Matsubara, S., L. A. Märtin, L., Merchant, A., Metzner, R., Müller-Linow, M., Nagel, K., Pieruschka, R., Pinto, F., Kuchendorf, C., Temperton, V., Thorpe, M., Dusschoten, D., Van Volkenburgh, E., Schurr, U., 2013. Non-invasive approaches for phenotyping of enhanced performance traits in bean. Functional Plant Biology 38, 968–983.

Rijsbergen, C.J.V., 1979. Information Retrieval. 2nd ed., Butterworth-Heinemann, Newton, MA, USA.

Rousseau, C., Belin, E., Bove, E., Rousseau, D., Fabre, F., Berruyer, R., Guillaumès, J., Manceau, C., Jacques, M.A., Boureau, T., . High throughput quantitative phenotyping of plant resistance using chlorophyll fluorescence image analysis 9, 17.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nature Methods 9, 676–682.

Smyth, D.R., 1995. Flower development: Origin of the cauliflower. Current Biology 5, 361 – 363.

Tremellat, V., 2017. Limiter l'impact des aléas climatiques sur chou-fleur via la création variétale et le développement d'outils de phénotypage. Technical Report. OBS Innovation.

van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T., the scikit-image contributors, 2014. scikit-image: image processing in Python. PeerJ 2, e453.

Watts, L.E., 1966. Investigations on the inheritance and responses to selection of riceyness in early summer cauliflower. Euphytica 15, 90–98.

Windt, C.W., Blümler, P., 2015. A portable NMR sensor to measure dynamic changes in the amount of water in living stems or fruit and its potential to measure sap flow. Tree Physiology 35, 366–375.

Wurr, D., Fellows, J., 2000. Temperature influences on the plant development of different maturity types of cauliflower. Proc. III Int. Symp. on Brassicas , 69–74.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks?. In: Advances in Neural Information Processing Systems, pp. 3320–3328.

Zhang, T.Y., Suen, C.Y., 1984. A fast parallel algorithm for thinning digital patterns. Commun. ACM 27, 236–239.

Zhou, Z.H., 2012. Ensemble Methods: Foundations and Algorithms. 1st ed., Chapman & Hall.

Zunair, H., Rahman, A., Mohammed, N., Cohen, J.P., 2020. Uniformizing techniques to process CT scans with 3D CNNs for tuberculosis prediction, In: Rekik, I., Adeli, E., Park, S.H., Valdés Hernández, M. del C. (Eds.), Predictive Intelligence in Medicine, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 156–168.

Table 1: List of features extracted from image of contour or image of skeleton and corresponding line colour used in Figures 7 and 8.

Circle diameter stands for diameter of the circle with same area of the category of the line;

Center distance stands for the category's center to contour center;

Area ratio stands for the category's area to contour area.

Features	**1**	**2**	**3**	**4**	**5**	**Line colour**
Contour	area	perimeter		circle diameter		green
Rectangle	area	perimeter	center distance	width to length	area ratio	blue
Hull	area	perimeter	center distance	circle diameter	area ratio	
Ellipse	area	perimeter	center distance	angle orientation	area ratio	red
Intensity	max	mean	min			

Table 2: The performance of classifiers based on features from image of contour or/and skeleton by Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and Partial Least Squares Discriminant Analysis (PLSDA) evaluated by F1 score and Matthews Correlation Coefficient (MCC). It is calculated separately on samples at stage 1, 2, 3, 4 or at mixed stage on 234; on samples of the year 2018 and 2019. All the results reaching expectation are marked in italics, with F1 more than 85% and MCC more than 65% (Boughorbel et al., 2017). One or 2 classifiers with best performance on every stage are marked with an asterisk. The slice distribution between H and S groups on every stage is also listed.

Data	**Stage**	**Features**	**Size**	**H**	**S**	**LDA F1**	**LDA MCC**	**QDA F1**	**QDA MCC**	**PLSDA F1**	**PLSDA MCC**
2018 + 2019	1	image of contour	88	22	66	47.37%	34.02%	47.06%	38.24%	84.89%	29.66%
2018 + 2019	1	image of skeleton	88	22	66	57.14%	49.92%	48.15%	27.28%	89.36%*	49.92%*
2018 + 2019	1	image of contour + skeleton	88	22	66	57.14%	43.84%	13.79%	2.42%	86.13%	38.22%
2018 + 2019	2	image of contour	54	34	20	_92.75%_*	_80.00%_*	84.85%	61.28%	_95.00%_*	_92.05%_*
2018 + 2019	2	image of skeleton	54	34	20	66.67%	0.62%	70.89%	-3.43%	36.84%	2.71%
2018 + 2019	2	image of contour + skeleton	54	34	20	83.78%	50.88%	81.16%	47.88%	73.68%	59.66%
2018 + 2019	3	image of contour	26	20	6	_92.68%_*	_65.92%_*	84.44%	-10.95%	54.54%	42.76%
2018 + 2019	3	image of skeleton	26	20	6	80.00%	13.33%	88.37%	37.36%	40.00%	27.24%
2018 + 2019	3	image of contour + skeleton	26	20	6	87.18%	49.08%	86.96%	0.00%	57.14%	42.60%
2018 + 2019	4	image of contour	152	96	56	_89.69%_*	_71.54%_*	_87.10%_	_67.04%_	74.55%	60.14%
2018 + 2019	4	image of skeleton	152	96	56	_89.01%_	_70.43%_	_90.26%_	_72.90%_	80.36%	68.90%
2018 + 2019	4	image of contour + skeleton	152	96	56	_91.10%_*	_76.06%_*	_90.82%_	_74.29%_	_85.71%_	_77.38%_
2018 + 2019	234	image of contour	232	150	82	87.30%*	62.64%*	86.01%	64.20%	78.21%	67.41%
2018 + 2019	234	image of skeleton	232	150	82	83.01%	50.22%	86.69%	58.28%	65.41%	47.46%
2018 + 2019	234	image of contour + skeleton	232	150	82	88.03%	64.46%	_88.67%_*	_67.93%_*	78.26%	66.74%
2018	4'	image of contour	136	80	56	_86.59%_	_66.37%_	83.54%	60.77%	80.00%	66.46%
2018	4'	image of skeleton	136	80	56	_88.75%_	_72.68%_	_90.68%_*	_77.18%_*	82.88%	71.09%
2018	4'	image of contour + skeleton	136	80	56	_88.34%_	_71.00%_	86.39%	64.87%	_90.57%_*	_84.94%_*
2018	234'	image of contour	172	94	78	84.38%*	64.73%*	79.10%	57.59%	77.92%	60.05%
2018	234'	image of skeleton	172	94	78	76.76%	49.76%	78.67%	47.72%	70.30%	43.32%
2018	234'	image of contour + skeleton	172	94	78	81.68%	58.85%	83.67%	62.42%	82.05%	67.16%

Figure 1: Examples of a healthy head versus stressed cauliflower heads with physiological disorders. The images were captured after they reach a diameter of around 10cm. A. Healthy head is tightly compact with only florets and forms one bracts; B. Open head has gaps among florets; C. Ricey head has protruding flower buds; D. Bracty head has leaves intermingled with florets.

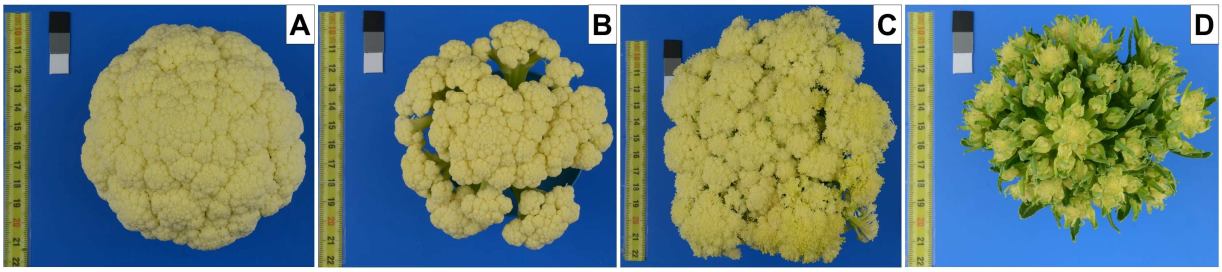


Figure 2: Example of a cauliflower plant at its primary mersitem stage. A,B. The whole plant has big leaves tightly wrapped around the apex; C. Once leaves are removed, the cauliflower apex becomes visible (circled in red); D,E. On the apex, floral primordia can be examined, here stained with carmine red for beter contrast.

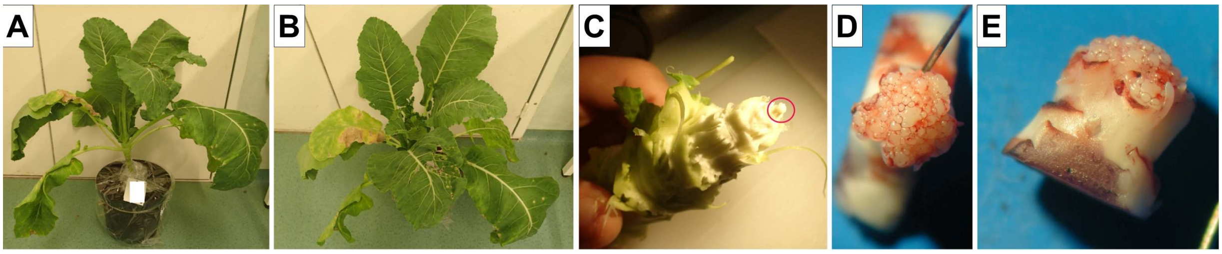


Figure 3: Example of one cauliflower plant placed in the "knee receiver coil" just before an MRI acquisition.

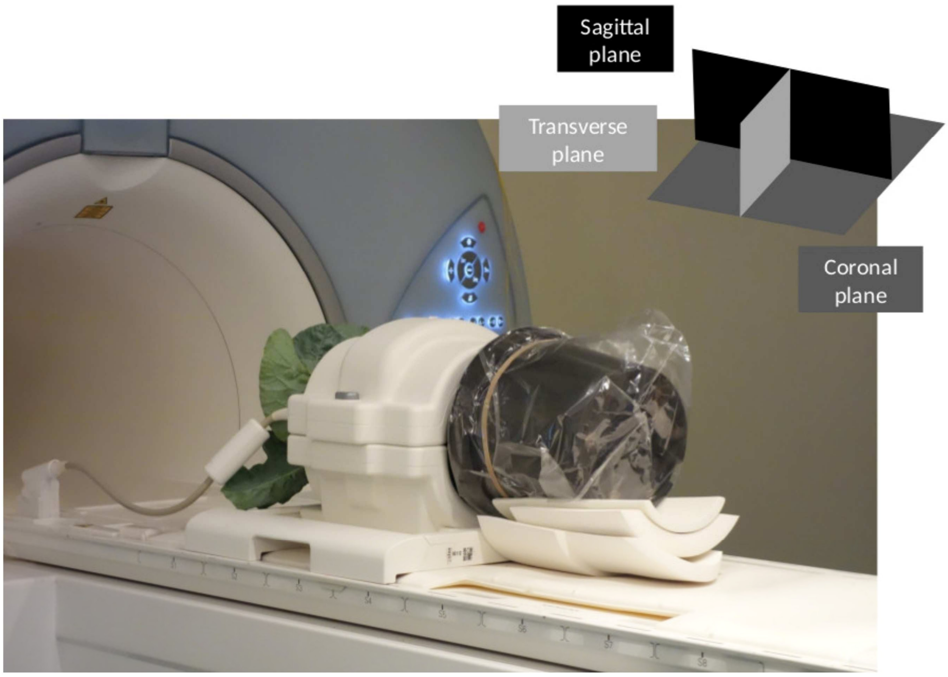


Figure 4: Floral induction at cauliflower apex illustrated by 4 developmental stages. Top line: schematic representation; middle line: examples in group H; bottom line: examples in group S. A,E,I: vegetative stage with only leaf scales (green/dark grey, plain line); B,F,J: curd-induction stage with enlargement of meristem center (blue/light grey, dashed line); C,G,K: curd-forming stage with round floral primordia (yellow/light grey, dashed line) initiated at the axil of each bract scale; D,H,L: curd-thickening stage with center only consisting of round floral primordia.

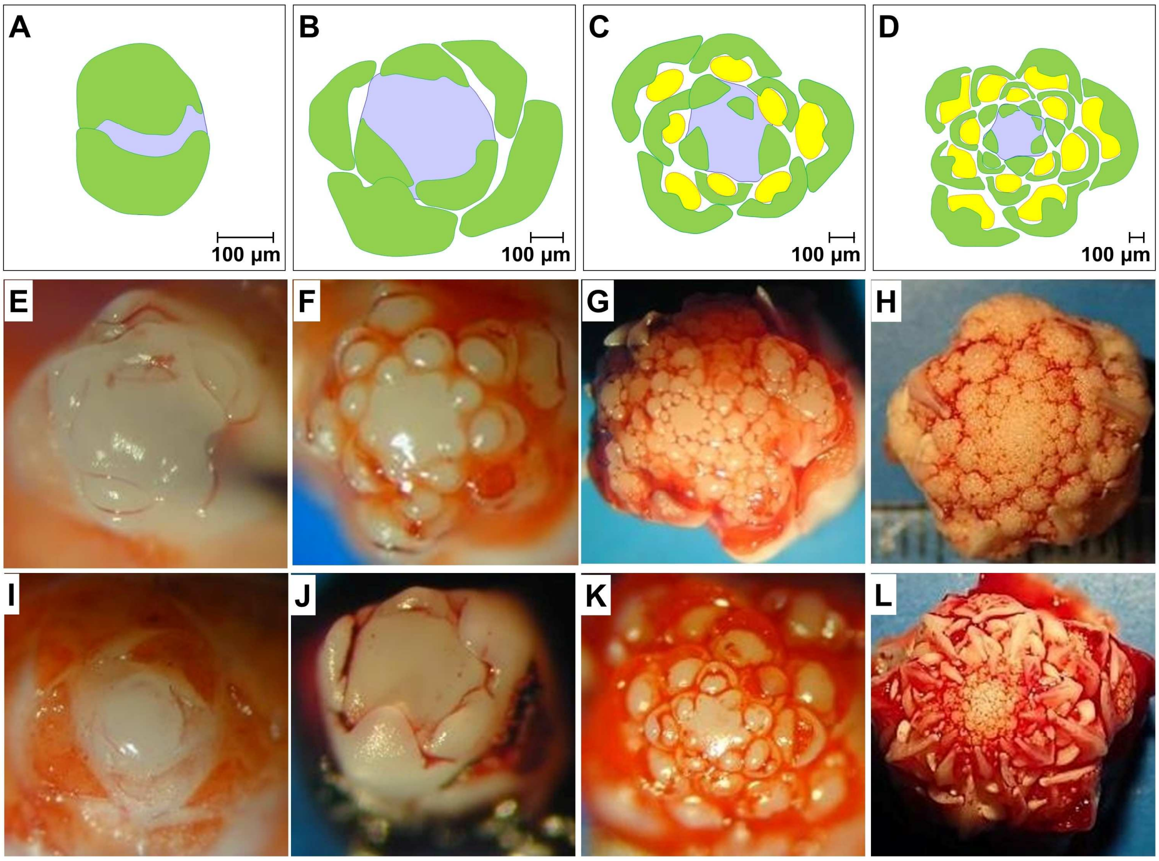


Figure 5: Examples of raw images in 2D planes and corresponding 3D reconstruction. A. raw image from a sample on plane XY; B. 3D reconstruction with this sample's raw images by Fiji [Schindelin et al. 2012]; C,D. middle slice on plane XZ and YZ. See Fig. 3 for a description of the planes.

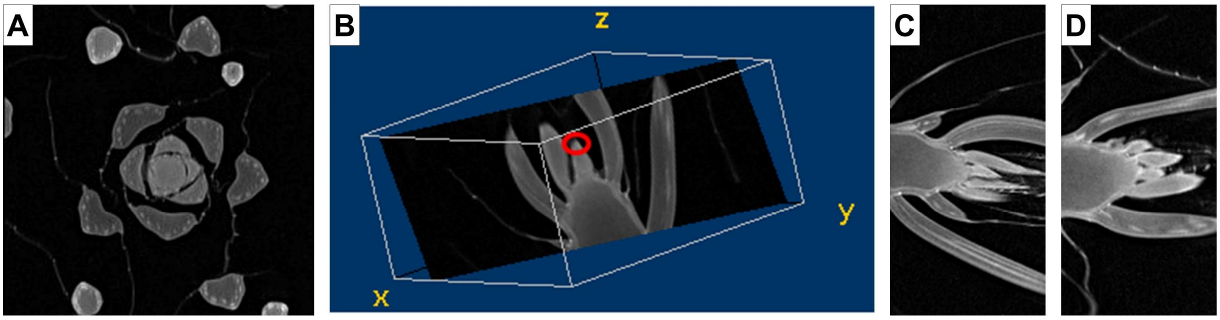


Figure 6: Illustrations for manual slice selection and histogram equalization A,B. missed meristem in middle slices of a sample on plane XZ and YZ; C,D. best illustrative slices of the same sample on plane XZ and YZ; E. best slice before histogram equalization; F. after Histogram Equalization.

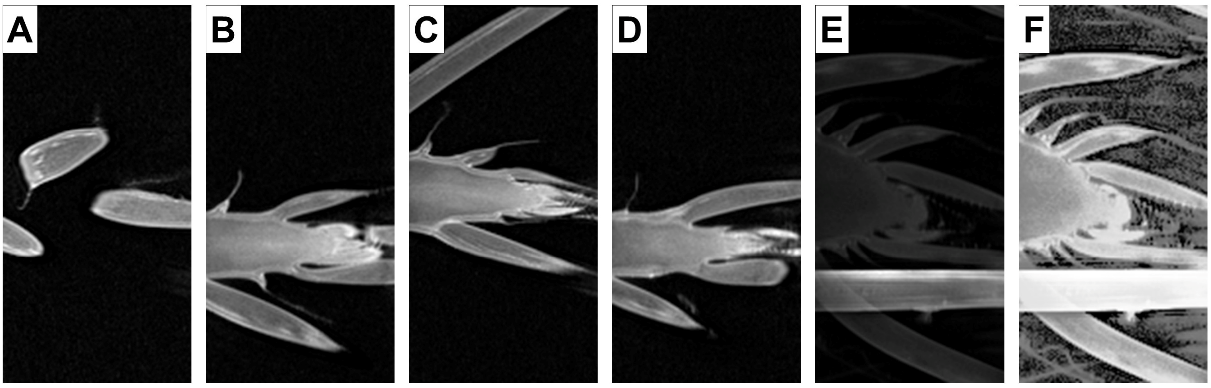


Figure 7: Extraction of the region of interest and resulting contour. A,B,C. illustration on extraction of region of interest. D. image of contour where contour is marked in green, rectangle in blue and ellipse in red.

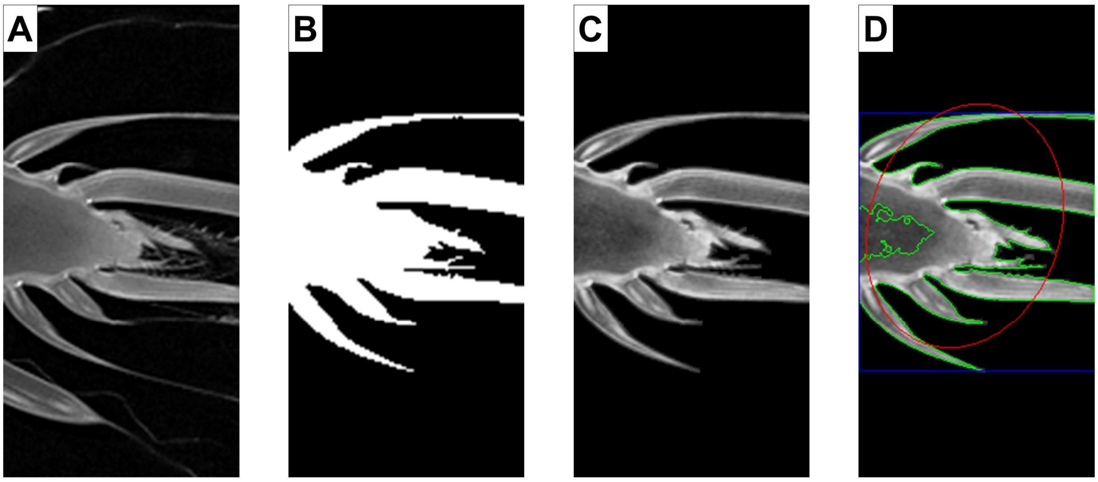
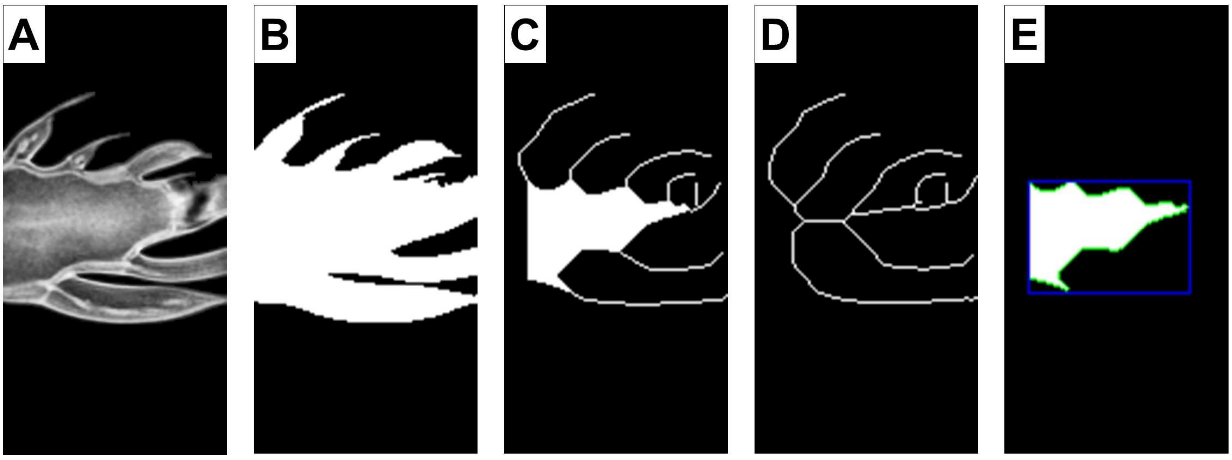


Figure 8: Extraction of the image of skeleton from a ROI image, from A. to E.



Appendix

Table 1: The Jaccard Index (JI) of classifiers based on features from image of contour or/and skeleton by Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Partial Least Squares Discriminant Analysis (PLSDA). For comparison, on stage 234, the 2D CNN JI was 64,26 % and the 3D CNN JI was 45.52 %.

Data	**Stage**	**Features**	**Size**	**H**	**S**	**LDA JI**	**QDA JI**	**PLSDA JI**
2018 + 2019	1	image of contour	88	22	66	77.27%	79.55%	76.14 %
2018 + 2019	1	image of skeleton	88	22	66	82.95%	67.05%	82.95 %
2018 + 2019	1	image of contour + skeleton	88	22	66	79.55%	71.59%	78.41 %
2018 + 2019	2	image of contour	54	34	20	90.74%	83.33%	96.30 %
2018 + 2019	2	image of skeleton	54	34	20	55.56%	57.41%	55.55 %
2018 + 2019	2	image of contour + skeleton	54	34	20	77.78%	75.93%	81.48 %
2018 + 2019	3	image of contour	26	20	6	88.46%	73.08%	80.77 %
2018 + 2019	3	image of skeleton	26	20	6	69.23%	80.77%	76.92 %
2018 + 2019	3	image of contour + skeleton	26	20	6	80.77%	76.92%	76.92 %
2018 + 2019	4	image of contour	152	96	56	86.84%	82.89%	81.58 %
2018 + 2019	4	image of skeleton	152	96	56	86.18%	87.50%	85.53 %
2018 + 2019	4	image of contour + skeleton	152	96	56	88.82%	88.16%	89.47 %
2018 + 2019	234	image of contour	232	150	82	83.19%	82.76%	85.34 %
2018 + 2019	234	image of skeleton	232	150	82	77.59%	79.74%	76.293 %
2018 + 2019	234	image of contour + skeleton	232	150	82	84.05%	85.34%	84.91 %
2018	4'	image of contour	136	80	56	86.03%	85.29%	83.82 %
2018	4'	image of skeleton	136	80	56	83.82%	86.03%	86.03 %
2018	4'	image of contour + skeleton	136	80	56	88.24%	85.29%	92.65 %
2018	234'	image of contour	172	94	78	82.56%	78.49%	80.23 %
2018	234'	image of skeleton	172	94	78	75.00%	73.26%	71.51 %
2018	234'	image of contour + skeleton	172	94	78	79.65%	81.98%	83.72%