Cotton Chronology: Convolutional Neural Network Enables Single-Plant Senescence Scoring with Temporal Drone Images

Aaron J. DeSalvio (aarondesalvio@tamu.edu)
Texas A&M University
Mustafa A. Arik (marik@tamu.edu)
Texas A&M University
Seth C. Murray (sethmurray@tamu.edu)
Texas A&M University
Oriana García-Ramos (orianagr123@gmail.com)
Texas A&M University
Serina M. DeSalvio (serinataluja@tamu.edu)
Texas A&M University
David M. Stelly (stelly@tamu.edu)
Texas A&M University

Method Article

Keywords: Deep learning, convolutional neural networks, drones, phenomics, cotton, senescence

DOI: https://doi.org/

Additional Declarations: No competing interests reported.

Cotton Chronology: Convolutional Neural Network Enables Single-Plant Senescence Scoring with Temporal Drone Images

Aaron J. DeSalvio[1], Mustafa A. Arik[2], Seth C. Murray[2*], Oriana García-Ramos[2], Serina M. DeSalvio[1], David M. Stelly[2]

[1]Interdisciplinary Graduate Program in Genetics and Genomics (Department of Biochemistry and Biophysics), Texas A&M University, College Station, TX 77843-2128, USA

[2]Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, 77843-2474, USA

*Author to whom correspondence should be addressed (sethmurray@tamu.edu).

Abstract

Senescence is a degenerative biological process that affects most organisms. Timing of senescence is critical for annual and perennial crops and is associated with yield and quality. Tracking time-series senescence data has previously required expert annotation and can be laborious for large-scale research. Here, a convolutional neural network (CNN) was trained on unoccupied aerial system (UAS, drone) images of individual plants of cotton (*Gossypium hirsutum* L.), an early application of single-plant analysis (SPA). Using images from 14 UAS flights capturing most of the senescence window, the CNN achieved 71.4% overall classification accuracy across six senescence categories, with class accuracies ranging between 46.8% to 89.4% despite large imbalances in numbers of images across classes. For

example, the number of images ranged from 109 to 1,129 for the lowest-performing class (80% senesced) to the highest-performing class (fully healthy). The results demonstrate that minimally pre-processed UAS images can enable translatable implementations of high-throughput phenotyping using deep learning methods. This has applications for understanding fundamental plant biology, monitoring orchards and other spaced plantings, plant breeding, and genetic research.

Keywords

Deep learning; convolutional neural networks; drones; phenomics; cotton; senescence

List of abbreviations

CNN: convolutional neural network

FHTP: field-based high-throughput phenotyping

ReLU: rectified linear unit

SPA: single-plant analysis

UAS: unoccupied aerial system, drone

Background

Senescence

Senescence is a term encompassing the summation of gene-, cell-, tissue-, and organism-level changes that lead to deterioration of biological

function. In monocarpic plants (mostly annuals), a single reproductive episode precedes the organism's death, whereas in polycarpic plants (mostly perennials), at least two reproductive episodes take place (1). Shuttling of assimilates from vegetative to reproductive organs in crop plants is a key feature of end-of-season senescence as it impacts harvest index of fruit, grain or seed, composition, as well efficiency of nutrient use (2). Animal and plant senescence can be conceptualized similarly, though due to the sessile nature of plants, senescence is specifically characterized by greater susceptibility to environmental stresses, pathogens, or physical damage (3). Contrary to appearing as a disorganized process at the cellular level, structural changes to senescing cells occur in an ordered manner, with leaf senescence under nuclear control (4, 5). Though leaf senescence is often viewed as a proxy for plant age, plant breeders frequently interpret it as a response to stress and factor it into selection decisions (6), underscoring importance and potential as a quantitative selection metric.

Despite history of cultivation dating back at least 3,000 years (7), *Gossypium* (cotton) maintains an indeterminate growth habit characteristic of perennials (8). In cotton, avoidance of adverse late-season weather conditions is dependent on optimal senescence timing, with premature senescence and late boll maturity potentially conferring reductions in fiber quality and yield (9). Leaf senescence can be accelerated by extreme high or low temperatures, with high temperatures promoting an increase in chloroplast reactive oxygen species, thus damaging the chloroplast, thylakoid

membrane, and photosynthesis-associated proteins, which ultimately impacts photosynthetic electron transfer (10). Drought, high salinity, and low temperatures can result in increased abscisic acid concentration in leaves, which is also associated with senescence progression (11). The source-sink ratio can impact aging of cotton, as well as cereal crops. Senescence of cotton was delayed by removal of fruiting branches (increasing the source-sink ratio) and accelerated by removing leaves (decreasing the ratio) (8, 12). Kumar et al. (13) proposed a model for source-sink-regulated senescence in maize in which programmed cell death and senescence are induced by the confluence of abscisic acid signaling, oxidative stress, and photosynthetic feedback inhibition.

High-throughput phenotyping analyses of senescence

Despite importance of senescence as a crop trait, robust evaluations of large numbers of genotypes and/or genotype-by-environment combinations are complicated by the need to evaluate across time based on repeated observations that span multiple stages of the overall maturation period. Because the flowering habit of cotton is indeterminate, the maturation period can last weeks to months. This “phenotyping bottleneck” for senescence might be effectively overcome using unoccupied aerial systems (UAS, also known as UAVs or drones) analysis and modeling of temporal phenotypes through the capture of images (14) which forms the basis of field-based high-throughput phenotyping (FHTP). Makanza et al. (6) previously reported a senescence index for a single time-point in maize derived from UAS images

and visual scores, both with moderately high heritabilities, with the senescence index displaying a significant association with grain yield. Using multispectral images of wheat, Hassan et al. (15) found temporal vegetation indices were more effective at selecting slow-senescing lines than a single ground-based score. Visually scoring senescence in time-series orthomosaics of maize hybrids, DeSalvio et al. (16) identified quantitative phenotypic indicators of senescence progression through plot-based temporal vegetation indices. While each of these studies demonstrated the applicability of UAS to decrypt senescence quantitatively using spectral data, further methodological development is required to increase the scale and throughput of phenomics-based senescence characterization for plant biology, plant breeding, genetics, and commercial applications.

Deep learning for plant phenotyping

Plant breeding programs depend on generational or yearly recording of phenotypic traits, many of which require time-consuming and labor-intensive collection methods. Accurately mapping connections between phenotype and genotype and ultimately saturating the phenome (17) will require analysis methods that can shuttle broad classes of data through automated processing pipelines requiring little modification in each use case. Representation learning circumvents the feature extraction required by conventional machine learning techniques by allowing algorithmic discovery of important features needed for classification or detection tasks (18). Beginning with raw inputs such as images, deep learning methods function

via a series of nonlinear layers that transform the raw input to slightly more abstract representations with each layer, ultimately amplifying signal from noise (18). For example, early layers may detect basic features such as plant material vs. soil while more abstract layers might enable distinction between different types of fungal disease within the same plant species. Within the deep learning class of models, convolutional neural networks (CNNs) are suitable for image recognition and categorization as they can learn complex and nonlinear mappings from large example data sets (19). CNNs are generally characterized by three types of neural layers: convolutional, pooling, and fully connected layers (20). During the forward stage of training, the input image is passed through each layer where the current weights and biases within each layer are applied, and the output (a prediction) is subsequently compared with the ground “truth” labels to calculate the loss. After each convolutional layer, nonlinearity is often introduced using the Rectified Linear Unit (ReLU) function (21) which also combats the vanishing gradient problem (22). During the second stage of training, backpropagation entails iterative application of the chain rule to calculate the gradient of the loss function with respect to each parameter, with parameters updated based on these calculations (20). Loss cost (differences between predicted and true labels) is reduced through repetitions of forward and backward stages until either a predesignated number of epochs is reached or a loss cutoff defining a stagnated learning rate is surpassed. Fully connected layers generally employ dropout to avoid overfitting (22-24).

CNNs have previously demonstrated utility in plant sciences, with applications including segmentation of overlapping field plants in maize (25), soybean stress (26), and disease detection in bell pepper, potato, and tomato (27), wheat (28), and within the PlantVillage data set, which includes 39 classes of plant leaves with varying diseases (29). Ubbens and Stavness (30) demonstrated an early application of neural networks for leaf counting, classifying mutants, and plant age using primarily the International Plant Phenotyping Network (IPPN) phenotyping data set (31). This was followed by the development of a method to augment plant phenotyping data sets using rendered synthetic plant images that improved model performance for counting leaves in rosette plants (32). Using UAS images, Hosseiny et al. (33) reported a CNN-based framework to count maize (*Zea mays* L.) plants from RGB images. Osco et al. (34) trained a CNN to both detect and geolocate plantation-rows in maize (*Zea mays* L.) and citrus (*Citrus sinensis*). These varied studies serve as early examples of a synergy between FHTP and deep learning. To our knowledge, no CNN method has yet been reported to enable single-plant analysis (SPA) of field-grown row crops. SPA would represent a paradigm shift from whole-plot analysis that is currently common in agricultural experiments. Most SPA studies to date have focused on individual tree phenotyping and have been conducted manually without the benefit of CNNs. Zarco-Tejada et al. (35) quantified tree height and crown shapes using digital surface models (DSMs) generated from UAS images; Díaz-Varela et al. (36) estimated olive tree crown parameters using DSMs;

Fujimoto et al. (37) developed a process to detect individual trees within a forest ecosystem, estimate forest structure, and predict future carbon dynamics. Early examples of SPA for row crops pre-dating the implementation of CNN-based analysis include an exploration of individual plant height variability in cabbage, pumpkin, barley, and wheat (38) and the analysis of the relationship between single-plant normalized difference vegetation index scores with full plot yield in winter wheat (*Triticum aestivum* L.) (39), however SPA remains underexplored for row crops. Novel SPA methods are hypothesized to enable early-generation selection in plant breeding, increased statistical power without requiring increased land usage, and refine dissection of genotype environment interactions at the single-plant level. From a crop improvement perspective, SPA could enable advancing the best individual plants to near homozygosity, saving time, space, resources, and removing the need to randomly select individual plants within a plot in single seed descent.

The methods proposed in this article serve to address the growing need for novel analysis methods in dissecting the plant phenome and conducting targeted plant biology studies as well as applied field breeding. The main objectives of this methods article were to: 1) develop a pipeline to rapidly segment orthomosaics from multiple time points into labeled individual-plant images for model training; 2) train a CNN to classify images of single cotton plants into six senescence categories; 3) demonstrate that time-intensive

preprocessing techniques such as cropping soil or removing background may not be needed to achieve robust classification accuracy.

Results

Across 235 cotton plants imaged 14 times, a total of 3,290 images were used, along with visual ratings classified from 0 (healthy) to 5 (senesced) to train the model. After 200 iterations of randomly splitting the data set by genotype into 80% training and 20% validation partitions (and further partitioning the training set into an 80/20 split) and training the model for 50 epochs in each iteration, the average classification accuracy (regardless of senescence class) was 0.714±0.030 (**Fig. 1A**). Compiling the training accuracy and validation accuracy revealed a plateau in validation accuracy near epoch 15, indicating the model could be trained with less time and computing resources while still achieving robust classification accuracy (**Fig. 1B**). In the same training scenario as previously described, loss converged to 1.506±0.396 (**Fig. 1C**). Near the 15-epoch mark, validation loss plateaued near 0.73 and started to increase past 20 epochs. A notable spread existed between loss and validation loss at epoch 50, which, as was observed with accuracy and validation accuracy, suggests early stoppage could be implemented for this model without detriment to classification accuracy (**Fig. 1D**).

To dissect the performance of the CNN within each of the six senescence classes, an average confusion matrix was calculated from

applying the model generated in each of the 200 iterations to the unseen data within the 20% partition of validation images. As described in Methods, images were minimally preprocessed: the dimensions were resized to be identical and the pixel values were normalized from [0, 255] to [0, 1]. Normalizing performance by predicted class revealed the model performed best in the senescence category of 5 at 89.4% accuracy, followed by predicting scores of 0 with 80.5% accuracy (**Fig. 1E**). The model was able to achieve moderate accuracy when classifying images with scores of 1 (63.3%) and 2 (58.0%), however in these cases, misclassification was between 14.0% and 22.6%, with the model often incorrectly predicting a lower score than the ground truth senescence label. The model struggled more with categories 3 (51.7%) and 4 (46.8%), with misclassification biased toward lower and higher scores in 3 and 4, respectively.

Precision, the number of correct predictions of a given class divided by all instances of that class in the data set, was highest for categories 0 and 5 (**Fig. 1F**). Precision decreased beginning with class 1 moving up to class 4, ranging from 0.655 to 0.381. For recall, the number of correct predictions of a given class divided by all predictions for that class, the CNN revealed the same order of performance, though categories 2, 3, and 4 revealed improvements over their respective precision values (**Fig. 1F**). The F1 score, the harmonic mean of precision and recall, was calculated for each category, and again revealed 5 and 0 as the top-performing categories with 1, 2, 3, and 4 again following in descending order (**Fig. 1F**).

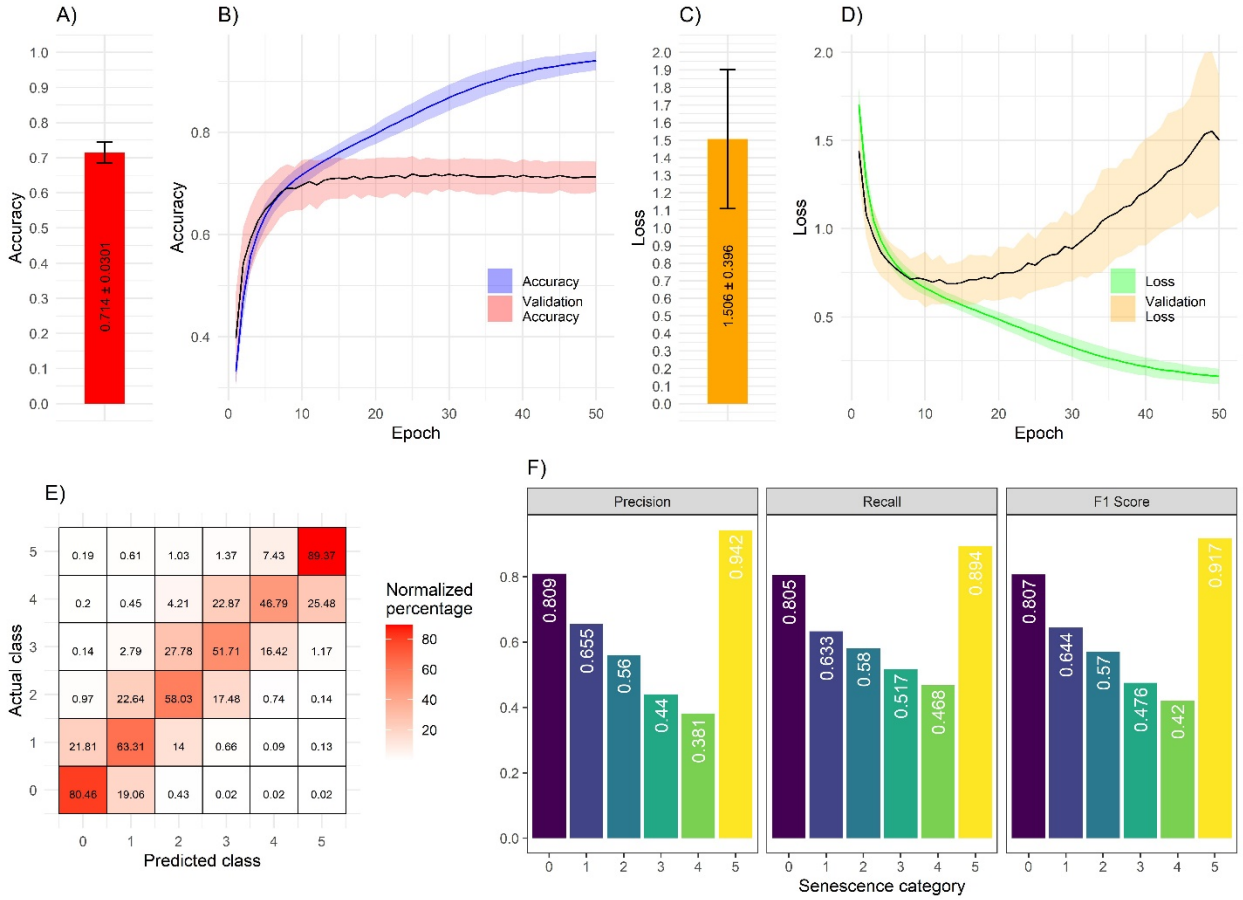

Figure 1. Evaluation metrics are reported for the six-category senescence scoring CNN. All metrics were calculated after training and evaluating the CNN in 200 instances, each with a random train/test split by genotype and each iteration allowed to train for 50 epochs. (**A**) Validation accuracy is given for the average classification performance, regardless of senescence category. (**B**) Accuracy and validation accuracy are reported with black lines indicating the average of each metric and shaded regions denoting ± standard deviation. (**C**) Validation loss across the entire training process. (**D**) Average loss and validation loss across 50 epochs. (**E**) The average confusion matrix normalized according to row. (**F**) Precision, recall, and F1 scores for each senescence category.

Discussion

In this article, a convolutional neural network (CNN) was trained to predict single-plant senescence scores using raw images extracted from RGB orthomosaics. This serves as the first incidence of applying field-based high-throughput phenotyping (FHTP) with drones to enable automated CNN single-plant senescence scoring. Notably, this model did not draw upon pre-

trained weights as an objective of the study was to evaluate the robustness of image-based learning with in-house data taken directly from temporal drone images. However, pre-trained networks have demonstrated success in tasks such as classifying tomato diseases (40), rice diseases (41), and identifying plant pests (42), though largely in controlled environments where SPA is routine. After developing a CNN aimed at identifying species of Miridae (“plant bugs”) using curated training images, Knyshov et al. (42) reported a model accuracy of 62% when applied to live field images, indicating generalizability of the model despite a modest number of images belonging to each class. In the present study, however, the model was only provided access to field images and achieved performance on par with this result, indicating that non-curated images have potential to enable robust classification accuracy with limited manual pre-processing before model training. A significant challenge in the present study was the uneven distribution of images belonging to each senescence score category (**Table 1**). Cotton, with wide spacing, is an ideal crop species to test single-plant analysis (SPA), however as a perennial, its ability to withstand drought and heat stress produced a disproportionate number of low-senescent plants with many plants exhibiting stay-green or demonstrating a resurgence in vigor (**Fig. 3**). Thus, it is expected that when trained on images of annual crops such as maize and sorghum where senescence must occur before harvest, the class imbalances will be resolved, providing more examples of scores in the 3 and 4 categories lacking in this data set; however, the close spacing and

overlapping canopies of such crops will pose novel challenges in FHTP SPA. Future SPA studies using cotton or other perennials might focus on introducing extreme heat or nutrient stress to induce senescence in a greater proportion of individual plants.

One of the most promising aspects of drone-based FHTP is its potential to assign quantitative measures to temporal phenotypes. As a selection metric, tracking senescence temporally can provide value to plant biologists and breeders seeking to elucidate relationships between different germplasm and critical agronomic traits such as heat stress tolerance, drought resilience, and timing of yield-related traits such as boll opening in cotton or grain filling duration in other crops. The CNN trained in this study provides early evidence that single-plant FHTP images can train models that are deeper than binary classification. Even with a limited number of example images of each category, a CNN could delineate differences between subtly distinct images. However, as demonstrated in **Fig. 1E** and **1F**, a higher number of examples of each class would likely have led to improved classification accuracy, precision, and recall. The ability of CNNs to filter signal from noise was demonstrated in this study as the lighting conditions between drone flights were often different, indicating the model was able to identify that plant pixels, not soil, were critical for delineating scores between images.

It is possible that the subjective visual “ground truth” ratings implemented in this study to score senescence were part of the classification

error, and future work might include an additional objective measure such as grouping/binning vegetation indices of individual plants across time into categories, thereby providing an objective measure of senescence scoring to be used as a comparison against visual rating. As in DeSalvio et al. (16), visual scores in this study were recorded based on orthomosaics, and using senescence-sensitive indices such as the red chromatic coordinate index may lead to improved model performance and could resolve some of the errors in the 3 and 4 categories seen in the confusion matrix (**Fig. 1E**).

The analysis pipeline developed in this study (**Fig. 2**) is adaptable to quantitative temporal analysis of single-plant images of other phenotypes involving spectral changes and could likely be applied to disease scoring or plant growth rates as estimated by vegetation indices. Though manual annotation (visual scoring or note-taking from plants or images of plants, or a field-based device) is required as a ground truth to train CNNs, the potential benefits to research programs from automating laborious and bias-prone plant selection metrics likely outweigh the time required for initial image annotation. A considerable amount of time within many research programs is spent in the data wrangling and curation phase before analysis is conducted, which delays decision-making regarding which genotypes are selected for future trials. CNNs have the potential to minimize or remove the need to annotate data with features such as spatial and environmental data, as the phenome of each plant, comprising the summation of all effects within and on a plant regarding genotype and environment, is of interest (17). CNNs can

learn these features from raw images and have the potential to supplement and improve visual selection methods that currently persist in plant sciences.

Conclusions

Using minimally pre-processed, time-series images of senescence in single cotton plants, a convolutional neural network (CNN) demonstrated the ability to classify senescence across six categories with 71.4% accuracy, with performance varying by category. The novel methods presented in this report highlight translatable implementations of field-based high throughput phenotyping (FHTP) and enable a paradigm shift from whole-plot analysis to single-plant analysis (SPA), which has the potential to allow more precise selection and advancement of germplasm, to quantify plant resilience at the individual level, and to study temporal morphology differences within genotypes.

Methods

Fig. 2 depicts a graphical summary of the methods employed in this article. Descriptions of methods used are listed in the order in which they appear.

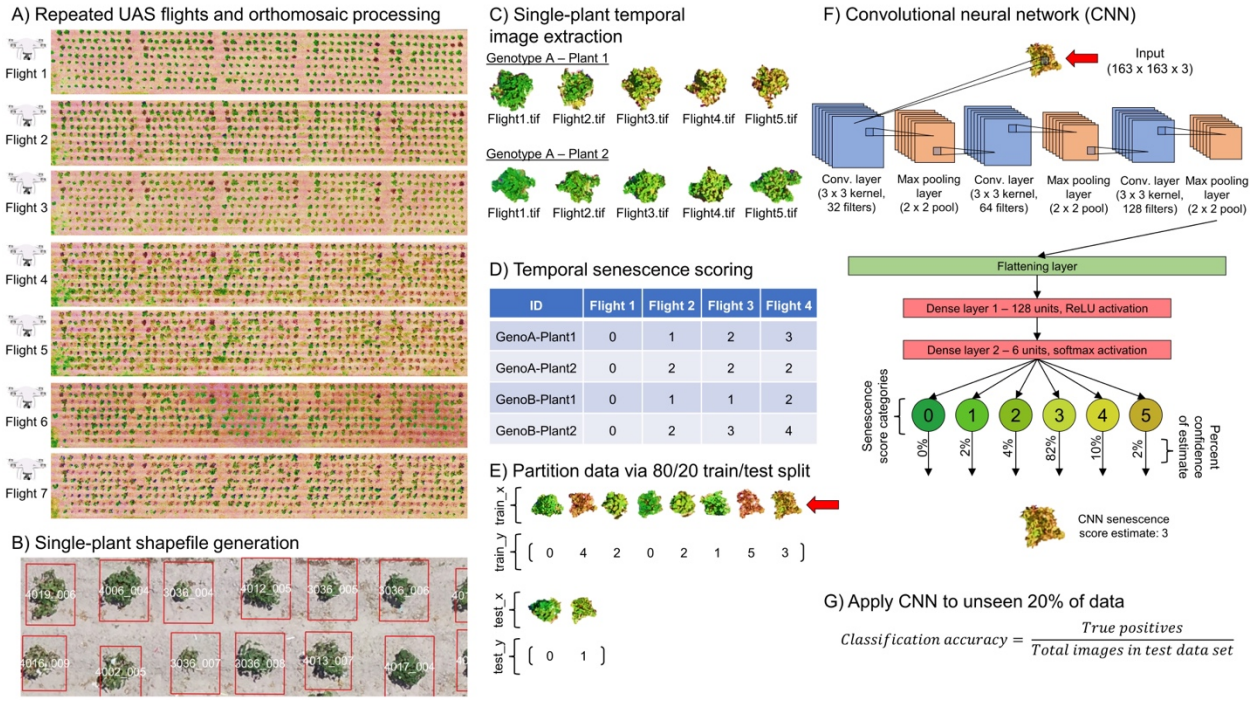

Figure 2. Graphical overview of data collection, image annotation, and model training. Except for (**B**), image saturation is enhanced to highlight differences in senescence progression between plants. Images were not modified in any way for CNN model training and testing purposes. (**A**) Orthomosaics from a representative sample of 7 of the 14 flights capturing the senescence window are shown in sequential order; (**B**) single-plant shapefiles were constructed and overlaid on each orthomosaic with minor adjustments to boundary boxes made as needed; (**C**) individual TIFs were extracted from each orthomosaic; (**D**) individual plants were scored for senescence; (**E**) data were partitioned by genotype into 80% (training) and 20% (testing) sections; (**F**) the CNN was constructed with three convolutional layers, three pooling layers, a flattening layer, and two dense layers; (**G**) model metrics were calculated including overall accuracy, recall, precision, and F1 scores for each class.

Cotton germplasm and experimental design

A field experiment was conducted in College Station, TX, between April and September 2023 to evaluate upland cotton (*G. hirsutum*) BC5 backcross-inbred lines containing small proportions of the genome of the wild Hawaiian cotton *G. tomentosum* (Nutt. ex Seem.), including 12 chromosome substitution lines, 35 chromosome segment substitution lines, two pure upland lines, a check, and a "filler". Greenhouse-grown three-week old

seedlings of 49 unique genotypes were mechanically space-transplanted into a randomized complete block design with 10 rows (*ca.* 200-cm spacing) of 10 plants (*ca.* 180-cm spacing), with outer rows and end hills serving as non-experimental "border", thus yielding 48 spaced transplants in each of the five blocks. Each genotype had approximately 5 replications with a total of 240 individual plants. Due to poor germination or other environmental causes, 5 plants perished early in the growing season, leaving 235 individual plants.

G. tomentosum is a wild allotetraploid species native to the Hawaiian Islands, where it is referred to as Ma'o (43). Several *G. tomentosum* traits are desirable for introgression into cultivated cotton species, including its characteristic heat tolerance, resistance to pests and diseases such as fleahoppers, tarnished plant bug, bollworm, and boll rot (44), thrips and jassids (45), and for desirable agronomic traits including fiber quality, length, and fineness (46, 47). Whole-genome sequencing and comparison revealed higher genetic similarity between *G. hirsutum* and *G. tomentosum* versus other wild allotetraploid relatives, such as *G. mustelinum* and *G. darwinii* (48).

High-throughput phenotyping of cotton fiber quality trial

Fig. 2A - After transplanting the individual plants to the field, UAS flights were conducted two or three times each week totaling to 46 flights across the growing season, of which 14 late-season flights are used to measure senescence. RGB images were captured with a DJI Phantom 4 Pro V2.0 with a 1-inch 20MP CMOS sensor with a mechanical shutter (SZ DJI

Technology Co., Ltd., Shenzhen, China). The DJI GS Pro application was used for mission planning, with all missions conducted at 20 m (above ground), 90% forward and 80% side overlap, flight speed of 1.0 m/s, and a 2.0 s shutter interval, producing a ground sampling distance of 5 mm/pix. Geotagged images were orthorectified and stitched with Agisoft Metashape Version 2.0.2 (Agisoft LLC, St. Petersburg, Russia) and orthomosaics (**Fig. 2A**) were generated via the program's structure from motion with multi-view stereo (SfM-MVS) workflow. Ground control points (GCPs) were recorded using an Emlid Reach M2 UAV RTK Kit (Emlid Tech Kft., Budapest, Hungary).

The procedures used to generate orthomosaics were as follows: 1) RGB images were imported into a Metashape project; 2) photos were aligned using referenced preselection, the key point limit was set to 40,000, the tie point limit was set to 4,000; 3) initial bundle adjustment was performed with the f, cx/cy, k1, k2, k3, p1, and p2 distortion parameters selected; 4) GCPs were imported as a .csv file and were manually tagged in six raw images per GCP; 5) all images were unchecked in the reference pane, the GCPs were integrated into the point cloud with the "update" button, and camera alignment was optimized using all available distortion parameters; 6) the dense point cloud and digital elevation map (DEM) were built using default settings; 7) the orthomosaic was constructed using the DEM as the surface.

Single-plant image extraction

Fig. 2B – Using the UAStools R package (49), a shapefile was generated in which a bounding box was placed over each plant. Boxes were

named according to the convention "*Genotype_ID*", where the genotype was a four-digit code corresponding to one of the 49 unique genotypes and ID was a three-digit code that delineated individual plants within the same genotype. Of the 46 flights conducted across the growing season, 14 were deemed as capturing the senescence window, occurring on: 24, 28, and 31 July; 4, 8, 11, 14, 16, 18, 21, 24, and 28 August; 1 and 5 September 2023. These orthomosaics were loaded into QGIS (50) along with the shapefile and manual corrections to bounding box dimensions and locations were made as necessary.

Single-plant temporal image extraction

Fig. 2C – The *st_read* function within the sf R package (51) was used to import the shapefile into the R environment. The *fieldCrop_grid* function from FIELDimageR.Extra (52) iterated through each bounding box within the shapefile and extracted one image per bounding box in TIF format (it is also possible to extract JPEGs), producing 240 images per flight. As five plants did not survive after being transplanted, effectively 235 images were obtained per flight, leading to a total data set size of 3,290 images (235 images per flight × 14 flights). Using a loop in R, images were renamed to the format "*YYYYMMDD-Genotype_ID.tif*" such that each image in the data set had a unique file name, where YYYYMMDD is the ISO 8601 date format of the drone flight. Images were separated into folders named according to the genotype and ID such that senescence could be scored rapidly in order of flight date for individual plants within each folder.

Senescence scoring

Fig. 2D – Within each folder (i.e., for each plant), 14 senescence scores were assigned in succession according to flight date and were recorded in tabular format (3,290 total senescence scores). A scoring system with six categories was implemented, where 0 = 0% senescence (completely green), 1 = 20% senescence, 2 = 40% senescence, 3 = 60% senescence, 4 = 80% senescence, and 5 = 100% senescence (completely brown). Examples of plants representing each score are shown in **Fig. 3**. Three distinct temporal phenotypes were observed: senescence progressed toward plant death (**Fig. 3A**), stay-green occurred and plants maintained vigor until the end of the season (**Fig. 3B**), or plants presented an initial drop in vigor but displayed resiliency and resurgence of vigor (**Fig. 3C**). The transient display of intermediate senescence stages led to an imbalanced data set, which was dominated primarily by scores of 0, 1, and 5, with notably less examples seen for 2, 3, and 4, respectively (**Table 1**).

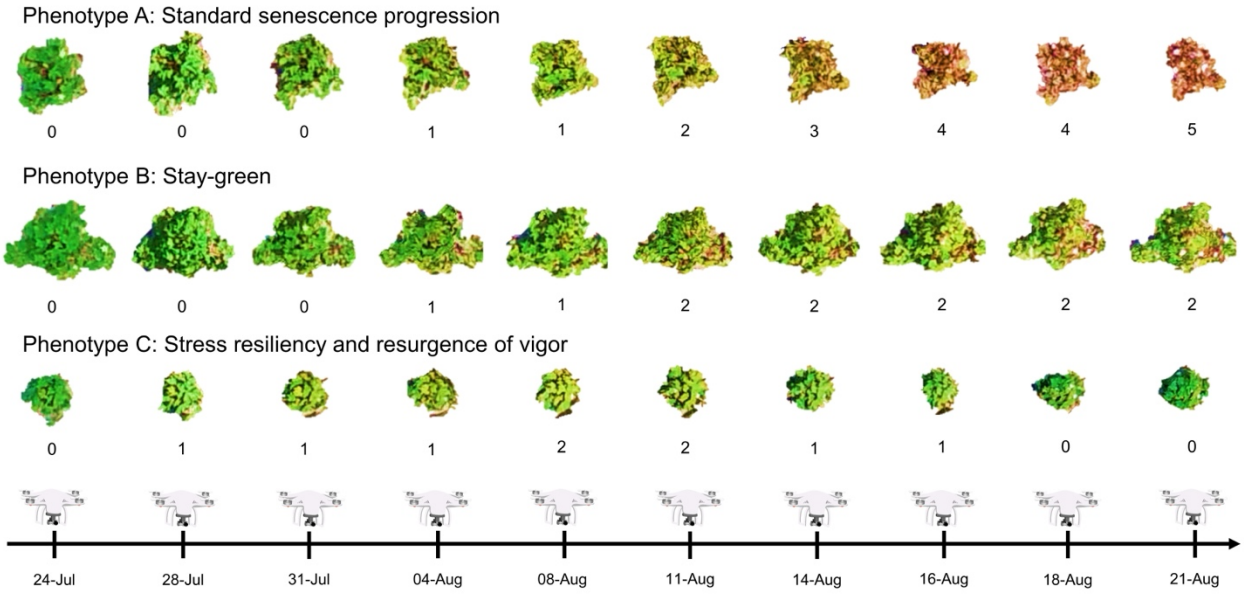

Figure 3. Three distinct temporal phenotypes of senescence are shown for individual plants across 10 of the 14 flights selected for senescence scoring. For “Phenotype A”, senescence progressed until permanent plant death. For “Phenotype B”, the stay-green phenomenon was observed, where the plant maintained intermediate greenness despite heat stress and drought. Lastly, “Phenotype C” experienced an initial senescent episode but recovered and displayed late-season vigor.

Table 1. Distribution of senescence scores belonging to each category. Since cotton is a perennial, many plants displayed either stay-green or a resurgence in vigor after an initial period of senescence, leading to a low number of scores of 3 (60% senescence) and 4 (80% senescence).

Score	0	1	2	3	4	5
Count	1,129	939	448	196	109	469

Data preparation and partitioning

Fig. 2E – All TIFs were imported in R via the *readImage* function within the EBImage package (53) and were resized such that all images had uniform dimensions of 163×163 pixels corresponding to the approximate average dimensions of each image. The *array_reshape* function from the reticulate R package was used to transform each image into a 163×163×3 array, thereby splitting the red, green, and blue color channels. A loop was constructed to successively run 200 iterations of training and evaluating the CNN. With each iteration, the data set was randomly split into 80% and 20% partitions based on genotype, with 188 genotypes (2,632 images) belonging to the training data set and 47 belonging to the testing data set (658 images). Training and testing image sets were each converted to 4D arrays, where the first dimension corresponded to the number of images (either 2,632 or 658), and the remaining dimensions corresponded to 163×163×3. Pixel values within

each 4D array were normalized to a range of [0, 1] from [0, 255] by dividing by 255. Vectors of ground truth senescence scores for training and testing images were one-hot encoded using the *to_categorical* function in the keras R package (54).

CNN parameters and model training

Fig. 2F – A CNN was parameterized using the *keras_model_sequential* function with the following settings:

1) The first 2D convolutional layer had 32 filters of dimensions 3□3 and used the ReLU activation function. A max-pooling layer follows with a 2□2 pool size.
2) The second and third 2D convolutional layers had 64 and 128 filters, respectively, both with 3□3 kernels that use the ReLU activation function. Each were followed by a max-pooling layer with a 2□2 pool size.
3) After the three convolutional layers, the output of the final max-pooling layer was flattened to transform the 2D feature maps into a 1D vector.
4) A dense (fully connected) layer with 128 units followed the flattening layer and used ReLU activation.
5) A dropout layer with a rate of 0.5 followed the first dense layer to prevent overfitting.
6) Another dense layer with 6 units served as the output layer, where each unit corresponded to one of the six senescence categories. A softmax

activation function was used in this layer to obtain probability values that sum to 1 for each prediction.

7) A categorical cross entropy loss function was used to compile the model. Root-mean-square propagation was used as the optimizer with a learning rate of 0.0001 and accuracy as the performance metric.

8) Using the *fit* function from keras, the model was trained for 50 epochs with a batch size of 32. The *validation_split* parameter was set to 0.2 to prevent overfitting. This resulted in the model being trained on 80% of the initial 80% training partition, or 64% of the original data set, resulting in approximately 2,106 training images.

Model evaluation

Fig. 2G – With each iteration of the CNN, the model was evaluated with the unseen set of images via the *evaluate* function in keras. The overall accuracy was calculated by the sum of the diagonal of the confusion matrix divided by the total number of testing images, given by $\frac{\text{True positives}}{658}$, with 658 being the number of validation images. For each iteration, the following metrics were saved: 1) a matrix of the accuracy, validation accuracy, loss, and validation loss for each epoch after subjecting the 20% validation data set to the model; 2) the average validation accuracy and loss across all epochs; 3) the confusion matrix. Both the confusion matrices and the averages of validation accuracy and loss were subsequently averaged across all 200 iterations. The average confusion matrix was used for calculations of precision, recall, and F1 scores. Where TP and FP denote true and false

positives, respectively, precision is given by $\frac{TP_{Class\ X}}{TP_{Class\ X} + FP_{Class\ X}}$ (correct class X predictions divided by all class X predictions), recall is given by $\frac{TP_{Class\ X}}{TP_{Class\ X} + FN_{Class\ X}}$ (correct class X predictions divided by all class X instances in the data set), and F1 score is the harmonic mean of precision and recall, given by $\frac{2 \times Precision \times Recall}{Precision+Recall}$. Precision, recall, and F1 scores were calculated for each senescence category.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All of the code used to assess the CNN, calculate evaluation metrics, and generate figures are available at the GitHub repository associated with this manuscript (55): https://github.com/ajdesalvio/cotton-chronology/tree/main. All files necessary to run the script, including the raw images, are available in the repository.

Competing Interests

The authors declare that they have no competing interests.

Funding

Financial support for this research has been provided by: USDA–NIFA–AFRI Award Nos. 2020-68013-32371, and 2021-67013-33915, USDA–NIFA Hatch funds, Texas A&M AgriLife Research, and the Eugene Butler Endowed Chair. AJD was supported by the National Science Foundation (NSF) Graduate Research Fellowship (GRFP). OGR and SMD were partially supported by Cotton Incorporated Awards 18-201 and 20-724, and NSF Award 1739092.

Authors' Contributions

Data curation and analysis were performed by AJD. MAA processed and extracted single-plant images. OGR, SMD, and DMS conceptualized and implemented the field experiment. AJD wrote the first draft of the manuscript, and MAA, SCM, OGR, SMD, and DMS each contributed to revising and editing of previous manuscript versions. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to acknowledge Dr. Robert Vaughn and the undergraduate students in the Cotton Cytogenetics Laboratory for their time and dedication in maintaining the field experiment.

NSF Statement

Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.

Authors and Affiliations

Interdisciplinary Graduate Program in Genetics and Genomics (Department of Biochemistry and Biophysics), Texas A&M University, College Station, TX 77843-2128, USA

Aaron J. DeSalvio (ORCID: 0000-0003-1818-4699), Serina M. DeSalvio (ORCID: 0000-0003-3741-7144)

Department of Soil and Crop Sciences, Texas A&M University, Agronomy Field Lab 110/111, College Station, TX, 77843, USA

Mustafa A. Arik, Seth C. Murray (ORCID: 0000-0002-2960-8226), Oriana García-Ramos (ORCID: 0000-0001-5399-7949), David M. Stelly (ORCID: 0000-0002-3468-4119)

References

1. Munné-Bosch S. Do perennials really senesce? Trends in plant science. 2008;13(5):216-20.

2. Gregersen PL, Culetic A, Boschian L, Krupinska K. Plant senescence and crop productivity. Plant molecular biology. 2013;82(6):603-22.

3. Leopold AC. Senescence in Plant Development: The death of plants or plant parts may be of positive ecological or physiological value. Science. 1961;134(3492):1727-32.
4. Gan S. Mitotic and postmitotic senescence in plants. Science of Aging Knowledge Environment. 2003;2003(38):re7-re.
5. Yoshida Y. Nuclear control of chloroplast activity in Elodea leaf cells. Protoplasma. 1962;54:476-92.
6. Makanza R, Zaman-Allah M, Cairns JE, Magorokosho C, Tarekegne A, Olsen M, Prasanna BM. High-throughput phenotyping of canopy cover and senescence in maize field trials using aerial digital canopy imaging. Remote Sensing. 2018;10(2):330.
7. Lee JA, Fang DD. Cotton as a world crop: origin, history, and current status. Cotton. 2015;57:1-23.
8. Chen Y, Dong H. Mechanisms and regulation of senescence and maturity performance in cotton. Field Crops Research. 2016;189:1-9.
9. Dong H, Li W, Tang W, Li Z, Zhang D, Niu Y. Yield, quality and leaf senescence of cotton grown at varying planting dates and plant densities in the Yellow River Valley of China. Field Crops Research. 2006;98(2-3):106-15.
10. Ougham H, Hörtensteiner S, Armstead I, Donnison I, King I, Thomas H, Mur L. The control of chlorophyll catabolism and the status of yellowing as a biomarker of leaf senescence. Plant Biology. 2008;10:4-14.

11. Lim PO, Kim HJ, Gil Nam H. Leaf senescence. Annu Rev Plant Biol. 2007;58:115-36.
12. Niu YH, Dong HZ, Li W-j, Li H-m. Effects of removal of early fruiting branches on yield, fiber quality and premature senescence in Bt transgenic cotton. Cotton Sci. 2007;19(1):52-6.
13. Kumar R, Brar MS, Kunduru B, Ackerman AJ, Yang Y, Luo F, et al. Genetic architecture of source-sink-regulated senescence in maize. Plant Physiology. 2023:kiad460.
14. Furbank RT, Tester M. Phenomics–technologies to relieve the phenotyping bottleneck. Trends in plant science. 2011;16(12):635-44.
15. Hassan MA, Yang M, Rasheed A, Tian X, Reynolds M, Xia X, et al. Quantifying senescence in bread wheat using multispectral imaging from an unmanned aerial vehicle and QTL mapping. Plant Physiology. 2021.
16. DeSalvio AJ, Adak A, Murray SC, Wilde SC, Isakeit T. Phenomic data-facilitated rust and senescence prediction in maize using machine learning algorithms. Scientific reports. 2022;12(1):1-14.
17. Murray SC, Adak A, Desalvio A, Lane H. Temporal field phenomics allows discovery of nature AND nurture, so can we saturate the phenome? Authorea Preprints. 2022.
18. LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015;521(7553):436-44.

19. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86(11):2278-324.

20. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: A review. Neurocomputing. 2016;187:27-48.

21. Glorot X, Bordes A, Bengio Y, editors. Deep sparse rectifier neural networks2011 2011: JMLR Workshop and Conference Proceedings.

22. Pawara P, Okafor E, Surinta O, Schomaker L, Wiering M, editors. Comparing local descriptors and bags of visual words to deep convolutional neural networks for plant recognition2017 2017: ICPRAM.

23. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research. 2014;15(1):1929-58.

24. Yoo H-J. Deep convolution neural networks in computer vision: a review. IEIE Transactions on Smart Processing & Computing. 2015;4(1):35-43.

25. Guo X, Qiu Y, Nettleton D, Schnable PS. High-Throughput Field Plant Phenotyping: A Self-Supervised Sequential CNN Method to Segment Overlapping Plants. Plant Phenomics. 2023;5:0052.

26. Ghosal S, Blystone D, Singh AK, Ganapathysubramanian B, Singh A, Sarkar S. An explainable deep machine vision framework for plant stress phenotyping. Proceedings of the National Academy of Sciences. 2018;115(18):4613-8.

27. Jung M, Song JS, Shin A-Y, Choi B, Go S, Kwon S-Y, et al. Construction of deep learning-based disease detection model in plants. Scientific Reports. 2023;13(1):7331.

28. Nigus EA, Taye GB, Girmaw DW, Salau AO. Development of a Model for Detection and Grading of Stem Rust in Wheat Using Deep Learning. Multimedia Tools and Applications. 2023:1-28.

29. Mohanty SP, Hughes DP, Salathé M. Using deep learning for image-based plant disease detection. Frontiers in plant science. 2016;7:1419.

30. Ubbens JR, Stavness I. Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. Frontiers in plant science. 2017;8:1190.

31. Minervini M, Abdelsamea MM, Tsaftaris SA. Image-based plant phenotyping with incremental learning and active contours. Ecological Informatics. 2014;23:35-48.

32. Ubbens J, Cieslak M, Prusinkiewicz P, Stavness I. The use of plant models in deep learning: an application to leaf counting in rosette plants. Plant methods. 2018;14:1-10.

33. Hosseiny B, Rastiveis H, Homayouni S. An Automated Framework for Plant Detection Based on Deep Simulated Learning from Drone Imagery. Remote Sensing. 2020;12(21):3521.

34. Osco LP, de Arruda MdS, Gonçalves DN, Dias A, Batistoti J, de Souza M, et al. A CNN approach to simultaneously count plants and detect

plantation-rows from UAV imagery. ISPRS Journal of Photogrammetry and Remote Sensing. 2021;174:1-17.

35. Zarco-Tejada PJ, Diaz-Varela R, Angileri V, Loudjani P. Tree height quantification using very high resolution imagery acquired from an unmanned aerial vehicle (UAV) and automatic 3D photo-reconstruction methods. European journal of agronomy. 2014;55:89-99.

36. Díaz-Varela RA, De la Rosa R, León L, Zarco-Tejada PJ. High-resolution airborne UAV imagery to assess olive tree crown parameters using 3D photo reconstruction: application in breeding trials. Remote Sensing. 2015;7(4):4213-32.

37. Fujimoto A, Haga C, Matsui T, Machimura T, Hayashi K, Sugita S, Takagi H. An end to end process development for UAV-SfM based forest monitoring: Individual tree detection, species classification and carbon dynamics simulation. Forests. 2019;10(8):680.

38. Jamil N, Kootstra G, Kooistra L. Evaluation of Individual Plant Growth Estimation in an Intercropping Field with UAV Imagery. Agriculture. 2022;12(1):102.

39. Crain J, Wang X, Evers B, Poland J. Evaluation of field-based single plant phenotyping for wheat breeding. The Plant Phenome Journal. 2022;5(1):e20045.

40. Aquil MAI, Ishak WHW. Evaluation of scratch and pre-trained convolutional neural networks for the classification of Tomato plant

diseases. IAES International Journal of Artificial Intelligence. 2021;10(2):467.

41. Shrivastava VK, Pradhan MK, Thakur MP, editors. Application of pre-trained deep convolutional neural networks for rice plant disease classification2021: IEEE.

42. Knyshov A, Hoang S, Weirauch C. Pretrained convolutional neural networks perform well in a challenging test case: identification of plant bugs (Hemiptera: Miridae) using a small number of training images. Insect Systematics and Diversity. 2021;5(2):3.

43. DeJoode DR, Wendel JF. Genetic diversity and origin of the Hawaiian Islands cotton, Gossypium tomentosum. American Journal of Botany. 1992;79(11):1311-9.

44. Saha S, Raska DA, Stelly DM. Upland Cotton (Gossypium hirsutum L.) x Hawaiian Cotton (G. tomentosum Nutt. Ex. Seem.) F1 hybrid hypoaneuploid chromosome substitution series. 2006.

45. Hulse-Kemp AM, Ashrafi H, Zheng X, Wang F, Hoegenauer KA, Maeda ABV, et al. Development and bin mapping of gene-associated interspecific SNPs for cotton (Gossypium hirsutum L.) introgression breeding efforts. BMC genomics. 2014;15(1):1-14.

46. Shim J, Mangat PK, Angeles-Shim RB. Natural variation in wild Gossypium species as a tool to broaden the genetic base of cultivated cotton. J Plant Sci Curr Res. 2018;2(005).

47. Zhang Z, Rong J, Waghmare VN, Chee PW, May OL, Wright RJ, et al. QTL alleles for improved fiber quality from a wild Hawaiian cotton, Gossypium tomentosum. Theoretical and applied genetics. 2011;123:1075-88.

48. Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, et al. Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. Nature genetics. 2020;52(5):525-33.

49. Anderson SL, Ii SCM. R/UAStools:: plotshpcreate: Create multi-polygon shapefiles for extraction of research plot scale agriculture remote sensing data. Frontiers in plant science. 2020;11.

50. QGIS Development Team. QGIS Geographic Information System. 2023.

51. Pebesma E, Bivand R. sf: simple features for R. R package version 0.6-3. URL: cran r-project org/package= sf. 2018;2.

52. Pawar PS, Matias FI. FIELDimageR. Extra: Advancing user experience and computational efficiency for analysis of orthomosaic from agricultural field trials. The Plant Phenome Journal. 2023;6(1):e20083.

53. Oles A, Pau G, Smith M, Sklyar O, Hube W. Image processing and analysis toolbox for R. R package version. 2020;4(0):1-52.

54. Allaire JJ, Chollet F. keras: R Interface to 'Keras'. R package version 2.3. 0.0. Computer software] https://CRAN R-project org/package= keras. 2020.

55. DeSalvio AJ. Supplementary Data - Cotton Chronology: Convolutional Neural Network Enables Single-Plant Senescence Scoring with Temporal Drone Images 2024 [Available from: https://github.com/ajdesalvio/cotton-chronology/tree/main.

Figures

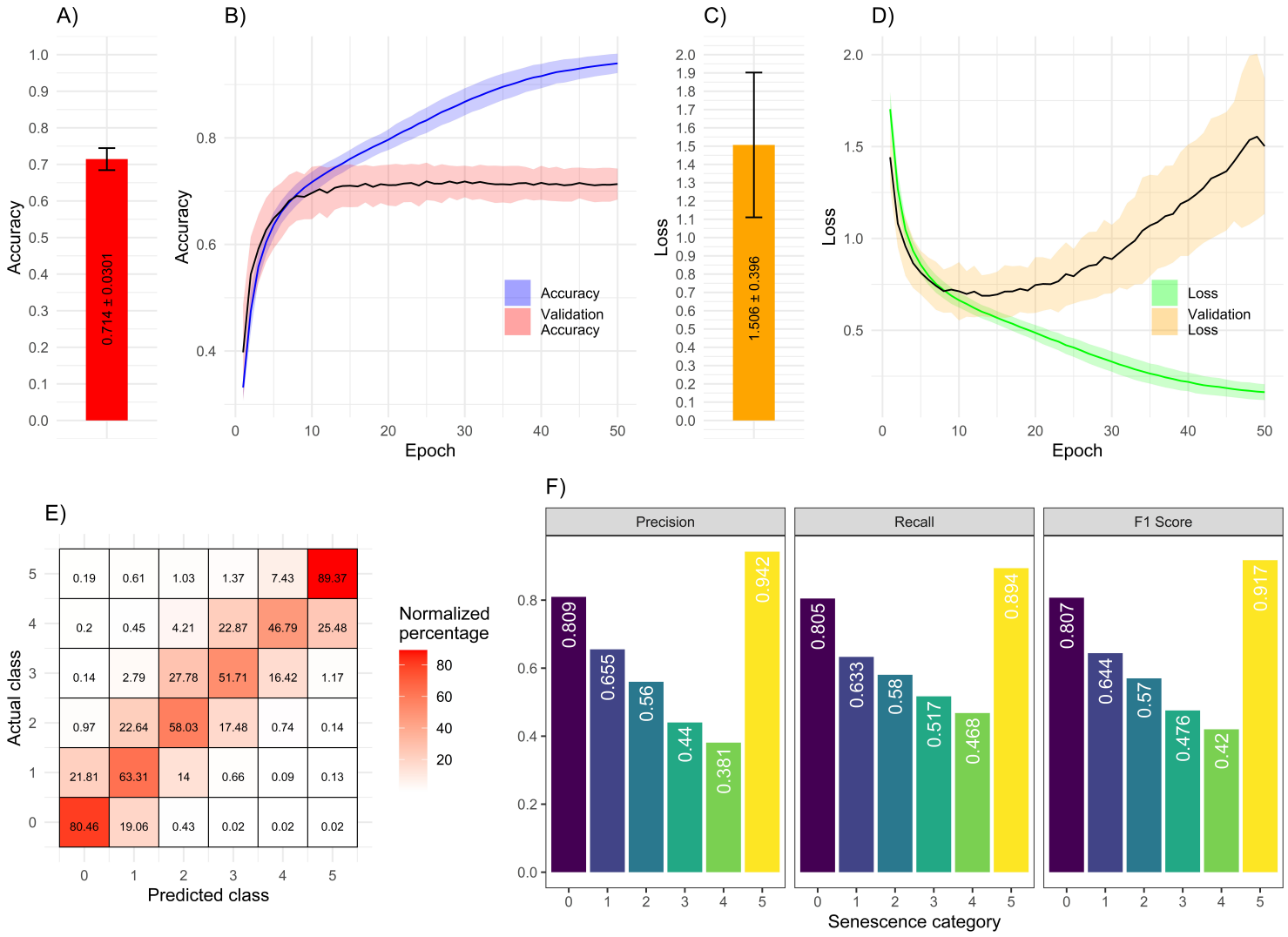

Figure 1

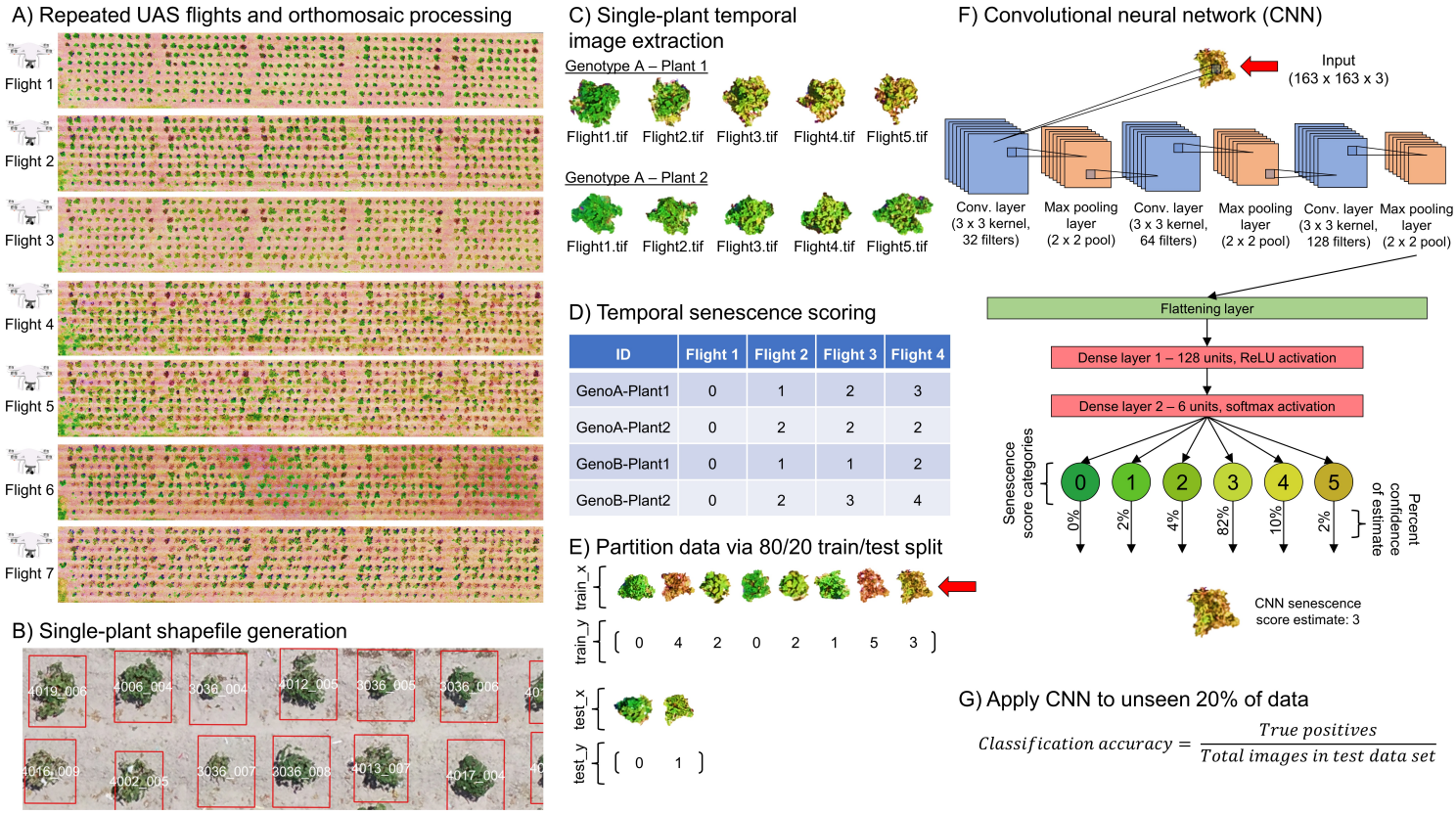

Figure 2

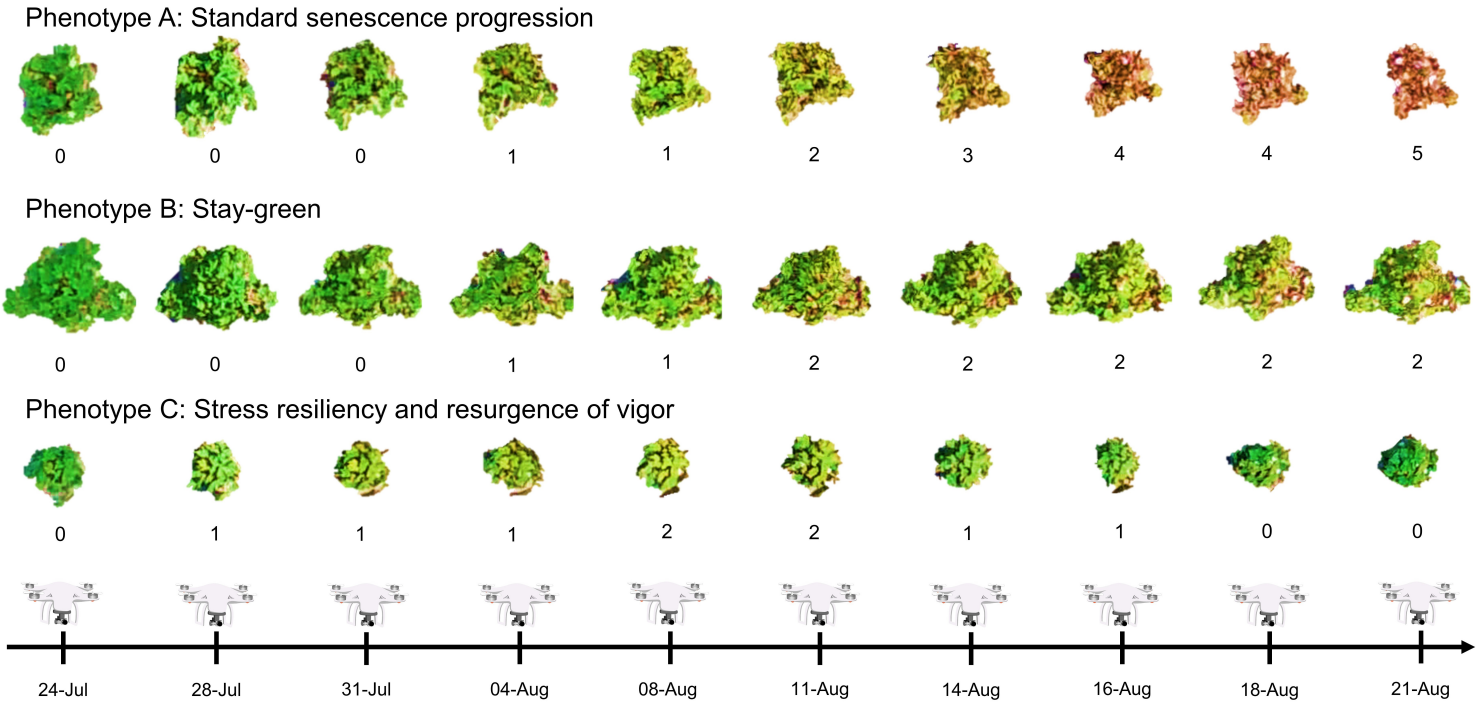

Figure 3