

Implementing Spatio-temporal 3D-Convolution Neural Networks and UAV Time Series Imagery to Better Predict Lodging Damage in Sorghum

Sebastian Varela ¹ and Andrew D. B. Leakey ^{1,2,3,4,5,*}

¹ Center for Advanced Bioenergy and Bioproducts Innovation, Urbana, IL 61801, USA; sv79@illinois.edu (S.V.)

² Institute for Genomic Biology, University of Illinois at Urbana Champaign, Urbana, IL 61801, USA

³ Department of Plant Biology, University of Illinois at Urbana Champaign, Urbana, IL 61801, USA

⁴ Center for Digital Agriculture, University of Illinois at Urbana Champaign, Urbana, IL 61801, USA

⁵ Department of Crop Sciences, University of Illinois at Urbana Champaign, Urbana, IL 61801, USA

* Correspondence: leakey@illinois.edu

Abstract: Unmanned aerial vehicle (UAV)-based remote sensing is gaining momentum in a variety of agricultural and environmental applications. Very high-resolution remote sensing image sets collected repeatedly across a crop growing season are becoming increasingly common. Analytical methods able to learn from both spatial and time dimensions of the data may allow improved estimation of crop traits, as well as the effects of genetics and the environment upon them. Multispectral and geometric time series imagery was collected by UAV on 11 dates, along with ground-truth data, in a field trial of 866 genetically diverse biomass sorghum accessions. We compared the performance of Convolution Neural Network (CNN) architectures that used image data from single dates (two spatial dimensions, 2D) versus multiple dates (two spatial dimensions + temporal dimension, 3D) to estimate lodging detection and severity. Lodging was detected with 3D-CNN analysis of time-series imagery with 0.88 accuracy, 0.92 precision, and 0.83 recall. This outperformed the best 2D-CNN on a single date with 0.85 accuracy, 0.84 precision, and 0.76 recall. Variation in lodging severity was estimated by the best 3D-CNN analysis with 9.4% mean absolute error (MAE), 11.9% root mean square error (RMSE), and goodness-of-fit (R^2) of 0.76. This was a significant improvement over the best 2D-CNN analysis with 11.84% MAE, 14.91% RMSE, and 0.63 R^2 . Success of the improved 3D-CNN analysis approach depended on inclusion of before and after data i.e. images collected on dates before and after the lodging event. Integration of geometric and spectral features with 3D-CNN architecture was also key to improved assessment of lodging severity, which is an important and difficult to assess phenomenon in bioenergy feedstocks such as biomass sorghum. This demonstrates that spatio-temporal CNN architectures based on UAV time series imagery have significant potential to enhance plant phenotyping capabilities in crop breeding and precision agriculture applications.

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Remote Sens.* **2022**, *14*, x. <https://doi.org/10.3390/xxxxx>

Academic Editor: Firstname Lastname

Received: date

Accepted: date

Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: 3D-convolution neural networks, time-series, lodging, sorghum.

1. Introduction

Lodging negatively affects yield in many crops [1–3]. This damage, defined as the displacement of plant stems from an upright position [4] is mainly induced by the strong winds of extreme weather events during the growing season of the crop. Lodging negatively impacts plant morphology [5], physiological processes [6], and growth [7], as well as impeding harvesting activities, to significantly reduce final above-ground yield of the crop [8–11]. Lodging damage can also introduce error or loss of phenotypic data used for selection in breeding, genetic analysis or management decisions [9, 10]. Due to its rapid vertical growth [12] and tall stature of up to 3–4 m, biomass-type *Sorghum bicolor* (L.) Moench is prone to lodging, especially in the second part of the growing season [13]. Therefore, the development of digital tools that provide precise and timely [14] evaluation of lodging would be of benefit to plant breeders, agronomists, and farmers.

Most studies on lodging have relied on in-situ visual inspection of the crop by domain experts i.e. agronomists or breeders [7] [15]. But, such methods require intensive manual labor and are limited to small spatial areas. Lodging has been monitored using satellites [14] and aircraft [16], including recently UAVs [3]. Advanced data-based modeling techniques benefit from increased resolution of spatial data collected by UAVs, as they are able to exploit this to learn relevant features on a given task [17]. Moreover, high quality data collection by UAVs is less constrained by cloud cover than when satellites or higher altitude aircraft are used, enabling more timely field inspection [4]. Multispectral imagery has been the most common source of information for lodging assessment using remote sensing imagery [4]. Under lodged conditions, the signal of different wavelengths that is reflected or backscattered to the sensor is altered by the changes in plant geometry and structure [2] [18]. Earlier work [19] successfully demonstrated the integration of spectral and textural features images to detect lodged areas of wheat with 82–91% accuracy. Other studies [20–22] have shown the value of integrating remote sensing imagery and machine learning to assess lodging damage. The first step in these approaches is a laborious manual search by domain experts for meaningful color and texture-based features, which are then marked to provide inputs for a classification algorithm to perform pixel-level mapping of lodging in certain areas [3]. While they represent important advances, there is potential to increase the accuracy and precision of the analysis, while also reducing labor. More recently, the ability of Convolution Neural Networks (CNNs) to maximize information gain from high-resolution UAV imagery has been demonstrated [23,24]. A CNN automatically generates representations of complex features, which provides an advantage over traditional machine learning and is suitable for automated image-based inference [24]. While CNNs have proved to be a powerful image analysis technique, most studies have only exploited variation along two dimensions in the data e.g. the two spatial dimensions of a single image (2D-CNNs) [25]. There is growing interest in adopting CNN models capable of exploiting a third dimension of variation e.g. along a time series (3D-CNNs) [26]. And, the recent exponential increase in the number of datasets that feature high-resolution imagery collected at frequent intervals over a growing season for a single field trial now make this a practical possibility. 3D-CNN architecture was used to predict end-of-season yield of wheat, barley, and oat at field scale by integrating UAV time series, weather, and yield monitor [17]. Sequential CNN-Long Short-Term Memory (LSTM) network architecture was used to predict soybean yield at county-scale by integrating MODIS satellite time series, weather, and national yield data [27]. In a different context, deep learning 3D-convolution-based architectures were utilized and benchmarked against traditional 2D-convolution architectures for hyperspectral imagery scene classification [27]. Together, these results across different use cases suggest that 3D CNN architectures can perform better than other architectures by exploiting their ability to use time as an additional descriptor of the target trait.

In this study, we propose to examine the effect of time, as an additional feature, on lodging damage prediction. Specifically, we focus on determining the value of persistent

aerial monitoring using spatio-temporal 3D-CNN time series models based on UAV time series imagery as input to determine both lodging detection and lodging severity.

The first objective is to assess the importance of the visual and near infra-red spectral and geometric features derived from the UAV imagery in the performance of the CNN models. Second, we test if 3D-CNN model architectures applied to time-series imagery can surpass the performance of 2D-CNN analysis of images from single dates. Third, we perform comparative evaluation of different time sequence configurations to understand the impact of the number of flights in the predictive sensitivity of 3D-CNN models. Fourth, we utilize activation maps to visualize and interpret CNN learning and predictive ability in these tasks.

2. Materials and Methods

2.1. Experimental site

A biomass sorghum experiment with an augmented incomplete block design of 1118 plots was planted on May 31, 2019, at the University of Illinois Energy Farm research facility, Urbana-Champaign (40.065789°N, -88.208477°W). The trial included (866) diverse accessions of biomass sorghum in single, four-row plots. In addition, six additional “check” accessions were planted in 16 four-row plots distributed across each of the 16 blocks, as has been previously described in detail [11]. A storm with strong wind produced lodging in plots across the experiment 82 days after planting (DAP), as shown in Figure 1. The ground-truthing survey was implemented immediately after the lodging event between 83-85 DAPs.

2.2. Lodging Assessment

We utilized a two-stage visual assessment of lodging to provide the data needed for model training and validation. The steps were to determine: 1) whether each plot was lodged or non-lodged; and 2) to estimate the severity of lodging, in the plots where it occurred, in terms of a lodging severity score initially suggested by [10] and adapted in [23]:

$$\text{Lodging score (1-100 \%)} = \text{Lodged area of plot (\%)} \times \text{Lodging severity}/3 \quad (1)$$

Lodging severity was defined as the estimated average inclination angle from vertical of plant stems in the lodged area of the plot. Three levels of lodging severity were defined ranging from mild (severity 1; 0-30°), to moderate (severity 2; 30-60°), and heavy (severity 3; 60-90°), as shown in Figure 1.

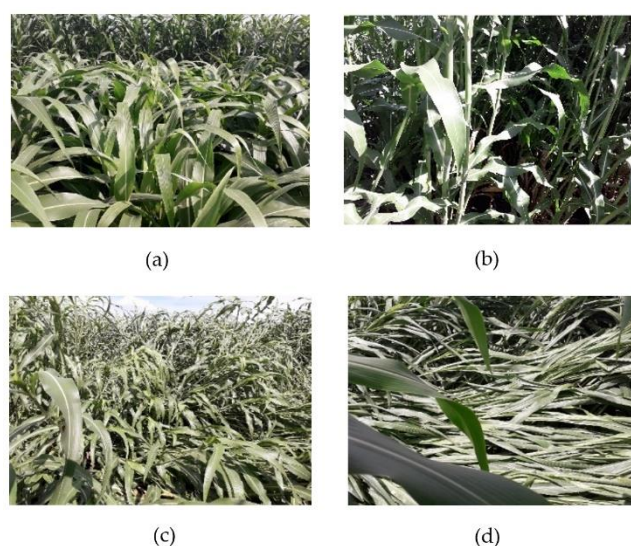


Figure 1. Example plots with different degrees of lodging severity during ground surveying. a) non-lodged, b) lodging severity 1 (mild), c) lodging severity 2 (moderate), d) lodging severity 3 (heavy).

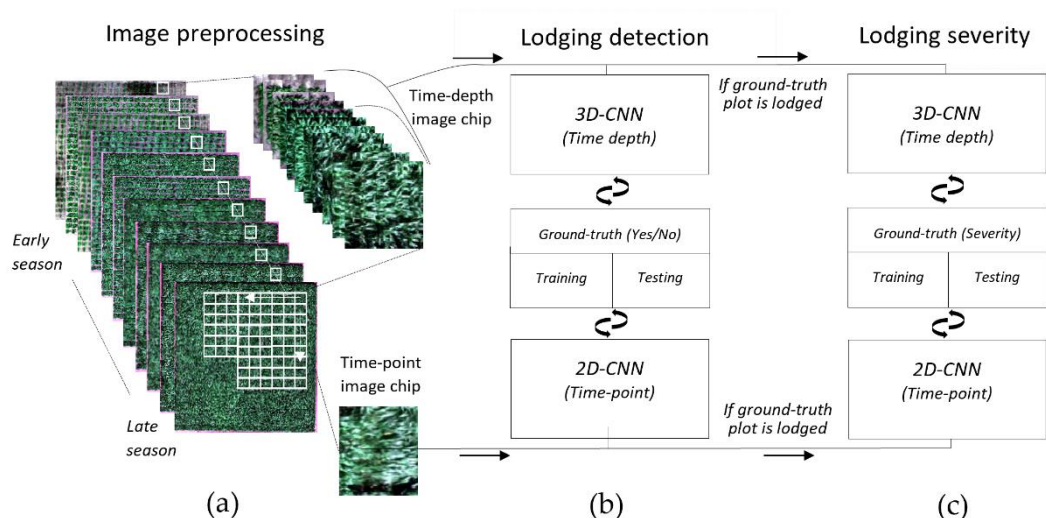
2.3 Aerial Data Collection and Preprocessing

A Rededge-M (Micasense, Seattle, WA, USA) multispectral sensor (MSI) was used for imaging to acquire both geometric and multispectral features. The camera includes five spectral bands in the blue (465 to 485 nm), green (550 to 570 nm), red (663 to 673 nm), rededge (712 to 722 nm), and near-infrared (820 to 860 nm) regions of the electromagnetic spectrum. The aerial platform utilized was a Matrice 600 Pro hexacopter (DJI, Shenzhen, China), equipped with a Gremsy T1 gimbal (Gremsy, Ho Chi Minh, Vietnam) used to mount the MSI sensor. Flights were conducted 11 times in the season between DAPs 31 and 97 under clear sky conditions and around solar noon. Flight altitude was set to 40 m above ground level, resulting in a ground sampling distance (GSD) of 2–3 cm/pixel. Flight planning considered a 90% forward and 80% side overlapping during image acquisition to ensure successful image postprocessing. Nine black and white square panels (70 cm x 70 cm) were distributed in the field and utilized as ground control points (GCPs). A RTK (real-time kinematic) survey was implemented using a Trimble R8 global navigation satellite system (GNSS) integrated with CORS-ILUC local station to survey the GCPs for accurate geo-referencing and co-registration of imagery through the season. A standard Micasense calibration panel was imaged on the ground before and after each flight for spectral calibration. Images were further imported into Metashape version 1.7.4 (Agisoft, St. Petersburg, Russia) to generate the two types of aerial image features: 1) multispectral, and 2) geometric 3D-reconstruction of the canopy known as crop surface models (CSMs) orthophotos.

An early season flight before plant emergence was utilized as ground level reference for the extraction of the absolute height in the CSM files on subsequent sampling dates ($n=11$). The resulting multispectral and CSM orthophotos were resampled to a common 5 cm/pixel resolution and stacked into one multi-feature and multi-date file object. Image chips for each plot were generated by clipping the orthophoto stack object using plot polygons shapefile corresponding to the experimental plot dimensions (2.8 m x 2.8 m) as shown in Figure 2. Three variations of the image chips were produced: (1) spectral features from RGB images (RGB); (2) spectral features from RGB, red-edge and NIR bands (MS); and (3) spectral and geometric features from RGB, red-edge and NIR bands along with a crop surface model (CSM_MS) (Table 1). These steps were implemented via *geopandas*, *shapely*, and *rioxarray* libraries in Python 3.7.11.

Table 1. Description of image features, number of flights, and image chips features dimension.

Model name	Image Features	Type of features	N features	N flights	Image chip size ⁴
RGB	RGB ¹	Spectral	3	1-11	48 × 48 × 3-33
MS	RGB, Re ² , NIR ²	Spectral	5	1-11	48 × 48 × 5-55
CSM_MS	RGB, Re, NIR, CSM ³	Spectral and geometric	6	1-11	48 × 48 × 6-66

¹ R=red, G = green, B=blue spectral bands.² Re = red edge, NIR=near-infrared spectral bands.³ CSM = Crop surface model.⁴ Image chip size = N pixels (width) × N pixels (height) × Depth (N flights × N features).**Figure 2.** Schematic workflow for lodging assessment. Orthophotos stacking, extraction of time-depth image chips (top) and time-point image chips (bottom) (a). Implementation of lodging detection (classification) prediction based on time-depth 3D-CNN and time-point 2D-CNN modelling and ground-truth data as plots labeled as lodged or non-lodged (b). Implementation of lodging scoring (regression) using time-depth 3D-CNN and time-point 2D-CNN modelling approaches and ground-truth data as lodging severity scoring (c).

2.4 CNN modelling

The core component of the CNN algorithm is the convolution operation, where a set of trainable kernels are applied to the input image to generate a set of spatial features that describes the target predictor [28]. The model learns basic features in the first layers and more complex features representations at deeper layers. The final output of a CNN is a set of features maps, depending on the use case, these can be directly utilized or fed to a fully connected layer either in classification or regression tasks. The 3D-CNN architecture is a variation of the traditional 2D architecture [29]. It utilizes all the standard CNN architecture features but adopts a third dimension, which in this case was the temporal descriptor of the target trait across the stack of images. The 3D convolution layer operation was applied with a learnable three-dimensional kernel considering the spatio-temporal characteristics of the image chip as width (number of pixels) × height (number of pixels) × depth (time sequence length).

2.5 CNN Implementation and Metrics

Customized CNN architectures were utilized in this study. After trial and error, for lodging detection, the backbone feature generator of the 2D-CNN architecture included

two 2d-convolution layers and one flatten layer (Figure 3a). Meanwhile for lodging severity, it considered three 2d-convolution layers and one flatten layer (Figure S1a). The rest of the architecture included a fully connected sigmoidal (classification) or linear (regression) layers heads, respectively. The 2D-convolution's kernel filter size was set to (3 pixels x 3 pixels x N features). The 3D-CNN architecture considered two 3d-convolution layers (Figures 3a and S1b). The 3D-convolution's kernel filter moves in three dimensions instead of two as the 2D-convolution filter, not only from left to right and from the top to the bottom, but also forward and backwards. Thus, its kernel's size was set to (3 x 3 x 2–9-time sequence lengths) operating at the sequence of each feature over time. The following settings were common to all CNN architectures; convolution layers set to 32 features, zero padding, stride equal to one with no overlapping, and rectified linear unit as activation function were considered (Figures 3 and S1). Batch normalization in the first convolution layer, 2D- max-pooling set to two following each convolution, with 3D- max-pooling equal to two over the spatial dimensions, equal to one over the depth dimension, and equal to two following the last convolution operation of the 3D- architecture, and 50 percent features dropout prior to the flatten layer step was also adopted.

2D-CNN model analysis was done on data from individual dates either before or after the lodging event. This tested the importance of the timing between the lodging event, ground-truthing, and aerial data collection. The following four dates of flights were tested:

- DAP 75; 7 days prior to lodging, labeled as '75 (1)'.
- DAP 83; 1 day after lodging, labeled as '83 (1)'.
- DAP 88; 6 days after lodging, labeled as '88 (1)'.
- DAP 97, 15 days after lodging, labeled as '97 (1)'.

where the values of 1 in parentheses indicates the analysis of data from a single date.

In addition, various configurations of input data from sequences of dates were analyzed using 3D-CNN architecture to test the sensitivity of predictions to variations in the length of the time sequence and the contribution of dates before and after the lodging event. The following six sequence configurations were built:

- DAPs 69,83; 13 days prior plus 1 day after lodging, labeled as '69_83 (2)'.
- DAPs 75,83; 7 days prior plus 1 day after lodging, labeled as '75_83 (2)'.
- DAPs 69,75,83; 13 and 7 days prior plus 1 day after lodging, labeled as '69:83 (3)'.
- DAPs 83,88,97; 1, 6 and 15 days after lodging, labeled as '83:97 (3)'.
- DAPs 62,69,75,83; 20, 13 and 7 days prior plus 1 day after lodging, labeled as '62:83 (4)'.
- DAPs 31,38,42,49,55,62,69,75,83; all available dates prior to lodging plus 1 day after lodging, labeled as '31:83 (9)'.

where the values of 2 to 9 in parentheses indicates analysis incorporating data from between two and nine dates.

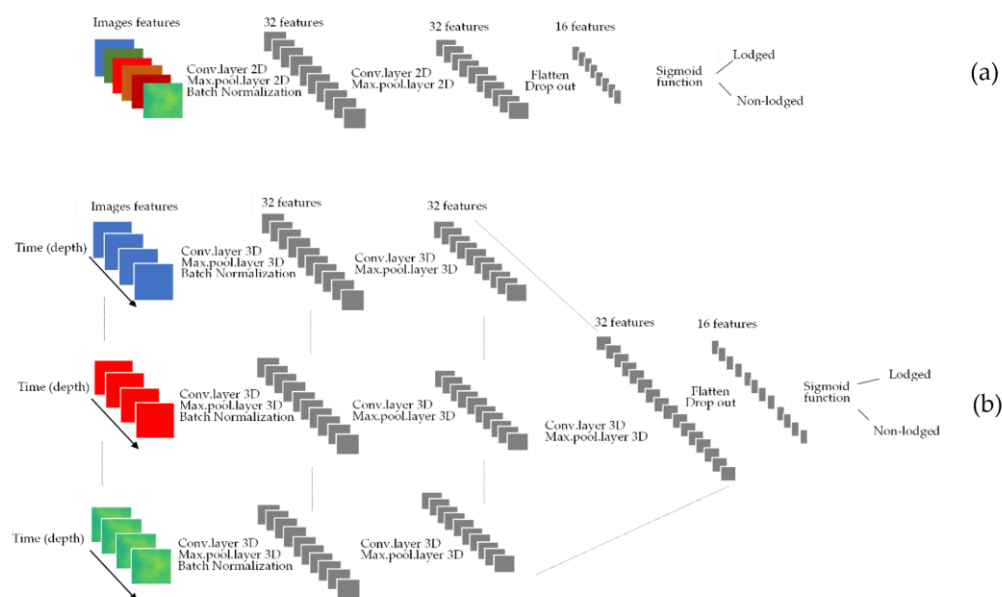


Figure 3. Diagram of custom CNN architectures utilized to assess lodging. 2D-CNN (a) and 3D-CNN (b) for lodging detection, CNN architectures for lodging severity can be found in supplemental materials Figures S1a and S1b.

CNN models were implemented in Python Keras API, Tensorflow-GPU version 2.7.0 using a NVIDIA GeForce RTX 3070 (8GB of GPU memory). The combination of different types of features, CNN architectures, and dates of flights or time sequence lengths resulted in the implementation of 60 models (Table S1). Half of these models were used to assess lodging detection, while the other half assessed lodging severity. In each of these tasks, 12 models were based on the 2D-CNN architecture as resulted from combining three types of features (Table 1) and the four dates of aerial data collection. The remaining 18 models were based on the 3D-CNN architecture, as the result of combining three types of features (Table 1) and six different configurations based on different lengths of time and the inclusion of data from dates before or after the lodging event. Each model fitting was iterated 10 times using a random training and testing partition to ease the convergence of models prediction metrics.

The full set of 1118 image chips were used to assess lodging detection, with 662 plots labeled as non-lodged, and 456 plots labeled as lodged during on-the-ground field surveying. The 456 plots labeled as lodged were further considered to assess lodging severity. In both tasks, training dataset was split 70:30 into training and testing datasets. The training dataset was split further 80:20 into training and validation datasets. The validation dataset was used to optimize models performance and prevent overfitting during training. Binary cross-entropy and mean square error were used as loss functions in each tasks, respectively. Models were trained up to 400 epochs using the early stop setting with lambda value 0.01 to ensure efficient training time. Learning rate and decay parameters were initialized with 0.01 and 0.001 values, respectively. The test dataset was utilized to expose the models to unseen data to evaluate the generalization ability of the models.

The overall accuracy (OA), precision, and recall were utilized as evaluation metrics to determine the performance of the models on lodging detection. The metrics are described in Equations 2-4:

$$OA = \frac{TP+TN}{TP+FN+FP+TN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4) \quad 283$$

In Equations (2), (3), and (4), TP (true positive) is defined as lodged plots correctly classified as lodged plots. TN (true negative) is defined as non-lodged plots correctly classified as non-lodged. FP (false positive) is defined as non-lodged plots incorrectly classified as lodged plots. FN (false negative) is defined as non-lodged plots incorrectly classified as lodged plots. 284
285
286
287
288

The mean absolute error (MAE), the relative root mean square error (RMSE) and coefficient of determination (R²) were utilized as evaluation metrics to determine the performance of the CNNs in lodging severity score. Each metric is described in Equations 5–6: 289
290
291
292

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5) \quad 293$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6) \quad 294$$

$$R^2 = 1 - [\sum_{i=1}^N (y_i - \hat{y}_i)^2 / \sum_{i=1}^N (y_i - \bar{y})^2] \quad (7) \quad 295$$

In Equations (5), (6), and (7), y , $\hat{y}$ and $\bar{y}$ are the observed, predicted, and observed mean lodging severity scores values of the i th plot, and n is the total number of samples within the study area. 296
297
298

Activation mapping visualization [30] was implemented to better understand the ability of the CNNs to use the information contained in the image arrays to deliver the output prediction. The importance of the image regions in the prediction can be identified by projecting back the weights of the output layer on to the convolutional feature maps, using this technique [31]. The following steps were used to generate activation maps in both tasks as described in [32]. First, the CNN model was utilized to map the input image to the activations of the last convolution layer as well as the output predictions. Then, the gradient of the predicted class (classification) or value (regression) for the input image with respect to the activations of the last convolution layer was computed. Lastly, each image channel in the feature map array was multiplied by how important this channel is with regard to the predicted class or value then sum all the channels to generate the corresponding activation map (Figure 6). 299
300
301
302
303
304
305
306
307
308
309
310

3. Results 311

3.1 Lodging detection 312

The overall accuracy (OA) of 2D-CNN predictions of the lodging that occurred during the storm 82 DAP was very limited when analyzing images collected seven days prior to the lodging event used as reference baseline (75 DAP; 0.61 to 0.73), increased substantially for images collected one day after the lodging event (83 DAP; 0.82 to 0.85), and then declined to moderate levels for images collected six to fifteen days after the lodging event (88 and 97 DAP; 0.73 to 0.75; Figure 4a). This variation in OA was associated with parallel changes in both precision (Figure 4c) and recall (Figure 4e). There were only modest effects on the OA of the 2D-CNN analysis when the number of image features was increased by adding multispectral or geometric information to RGB, except when analyzing pre-lodging images (75 DAP; Figure 4a). In that specific case, adding CSM data increased OA mainly by increasing precision (Figure 4c), but also through gains in recall (Figure 4e). 313
314
315
316
317
318
319
320
321
322
323
324

Applying 3D-CNN analysis to images from multiple sampling dates produced modest improvement in the OA of lodging detection compared to 2D-CNN analysis of data from single dates (Figure 4a). OA was greatest (0.86-0.88) when the 3D-CNN only analyzed imagery collected on only 2 to 3 dates that straddled the lodging event (Figure 4b). But, performance only declined marginally when images from more dates both before and after the lodging event were included, and results were especially robust when geometric features were combined with spectral features (CSM_MS) compared to RGB or multispectral (MS) features alone (Figure 4a). This was consistent with the greatest gains in performance of 3D-CNN over 2D-CNN occurring when using CSM_MS data compared with MS or RGB. The geometric features included in CSM_MS were particularly important in maintaining high levels of precision and recall, which decreased for 3D-CNN analysis of RGB and MS as the number of dates of data collection increased (Figures 4d-f).

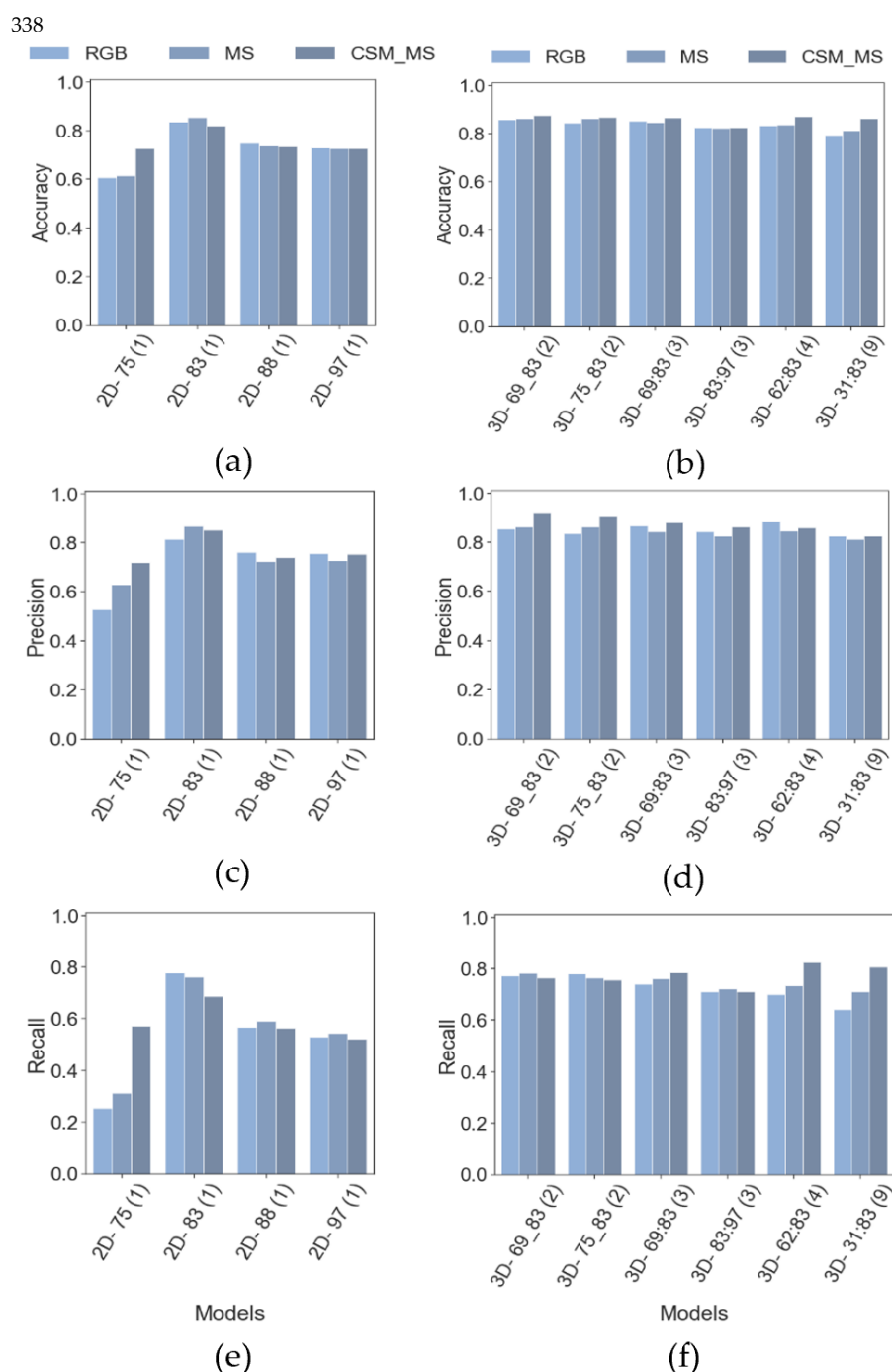


Figure 4. Evaluation of lodging detection prediction in testing data for time-point 2D by DAP and 3D- models by number of time frame sequences considered via OA (a, b), precision (c, d) and recall (e, f) metrics.

3.2 Lodging Severity

When predicting lodging severity, the patterns of variation in the performance of 3D-CNN and 2D-CNN analyses across the various data input scenarios mirrored those observed for lodging detection, but the benefits of 3D-CNN over 2D-CNN were greater (Figures 5a-f). This was specifically the case when the additional information found in the geometric and spectral features of CSM_MS data were leveraged over the progressively simpler image features of MS and then RGB data by the 3D-CNN (Figures 5b, d, f). The predictive ability of the 3D-CNN analyses of CSM_MS data was good (MAE = 9.44-

10.06%, RMSE = 11.88-12.62%, $R^2 = 0.72-0.76$) whenever imagery from dates before and after the lodging event (82 DAP) were used. Predictive ability declined only slightly from its maximum when images from only two dates were used or when images from across nine dates were used (Figures 5b, d, f). By comparison, when CSM_MS data from only the three dates following the lodging event were used, R^2 decreased and error increased (Figure 5a-f).

Despite being significantly less capable than the best 3D-CNN models, the 2D-CNN analysis of images from one day after the lodging event that occurred 82 DAP was able to predict lodging severity moderately well (MAE = 11.84%, RMSE = 14.90%, $R^2 = 0.63$; Figure 5a, c, e). However, the ability to predict lodging severity progressively declined using imagery that was collected six and then fifteen days after the lodging event (Figure 5a, c, e). And, almost no variance in lodging severity across the sorghum population could be explained by a 2D-CNN using images collected 7 days prior to the lodging event. As with 3D-CNN models, the extra spectral and geometric features in the CSM_MS dataset compared to the MS and RGB data aided improved predictions of lodging severity.

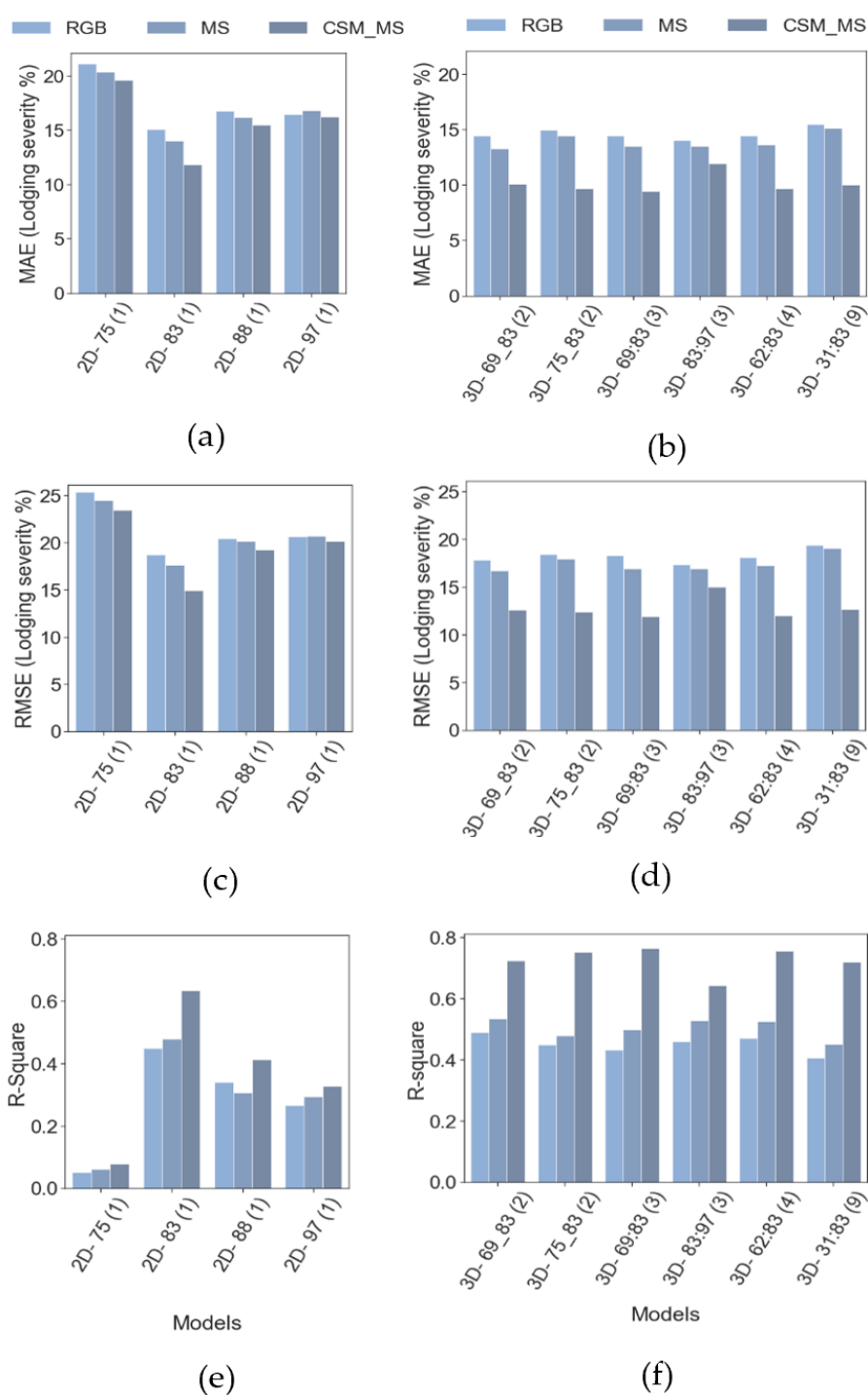


Figure 5. Evaluation of lodging severity prediction in testing data for time-point 2D-CNN by DAP and time-depth 3D-CNN models by number of time frame sequences considered via MAE (a, b), RMSE (c, d) and R2 (e, f) metrics.

3.3 Interpretation of CNNs predictions via activation mapping visualization

The superior performance of 3D-CNN analysis over 2D-CNN analysis can be visualized for the scenarios of each type with greatest predictive power (i.e. 2D-83 and 3D-69:83 for CSM_MS) using receiver operating characteristic (ROC) curves, observed versus predicted values for both lodging traits, and by interpretation of the activation maps (Figures 6a-e).

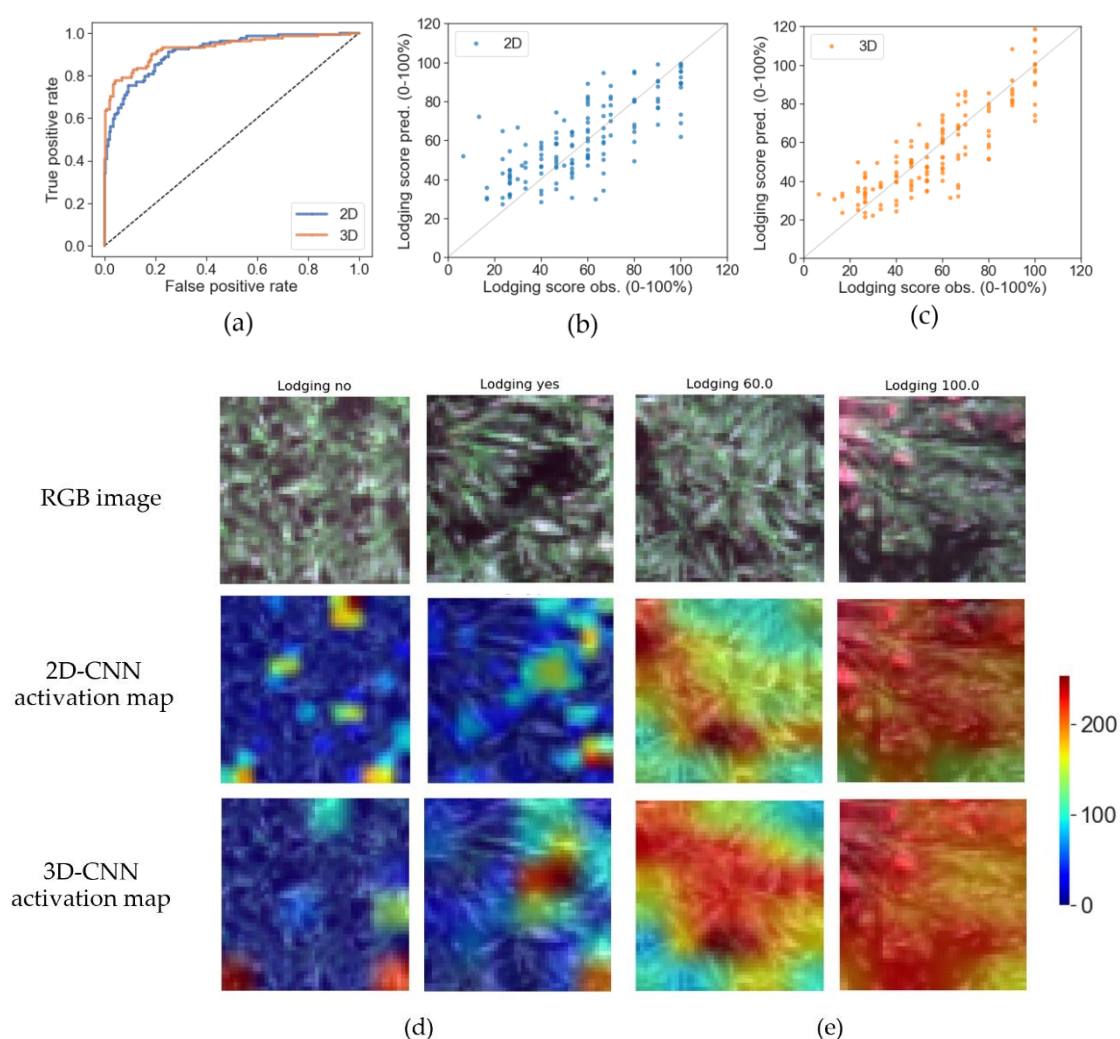


Figure 6. Evaluation of best 2D MS and 3D CSM_MS models on lodging detection prediction via ROC curves (a), and lodging severity observed and predicted (b) in testing dataset. Visual representation of activation maps [30] of last max pooling layer of each CNN for lodging detection (c), and lodging severity (d) assessment. Example RGB image of plots (top), corresponding gradient activation maps of best 2D-CNN (middle), and best 3D-CNN (bottom) in testing dataset. The level of activation of CNN layers in the images is represented by low (blue), intermediate (yellow), and high (red).

The ability of 3D-CNN analysis to classify lodged plots correctly (true positives) without incorrectly considering plots to be lodged (false positives) is clear in the ROC curves (Figure 6a). Similarly, 3D-CNN analysis predicted lodging severity with less error and bias than 2D-CNN models (Figures 6a, b). Both types of models tend to overestimate predicted values in plots where mild lodging occurred (0-30% of lodging severity), but this pattern was less of an issue for the 3D-CNN models.

Activation maps indicated that 2D-CNN and 3D-CNN models relied on similar regions within images for lodging detection or prediction of lodging severity (Figures 6d, e). Areas of high activation were visibly located in the portions of images where lodging was most severe (Figure 6d). However, it is notable that 3D-CNN models for both detection and severity estimation were able to activate a larger region around the area where plants were damaged. And in the case of severe lodging (100%), the 3D-CNN model tended to fine-tune the activation level in accordance with heterogeneity in lodging within the plot (Figure 6e).

4. Discussion

This study analyzed two spatial dimensions and a temporal dimension in data from a time-course of UAV remote sensing images with a 3D-CNN architecture to facilitate improvements in lodging detection and prediction of lodging severity relative to standard 2D-CNN architecture (Figures 4 and 5). Modest gains were made in the qualitative task of lodging detection through use of a 3D-CNN compared to a 2D-CNN, which had an overall accuracy of 0.79–0.88 depending on the number of spectral and geometric features of the input data. In contrast, when estimating quantitative variation in lodging severity, there could be significant benefits to using a 3D-CNN over a 2D-CNN. These gains in prediction ability were realized when a 3D-CNN was used on feature-rich data containing both geometric and spectral information i.e. when the CSM or red-edge and NIR wavebands were dropped from the input data, the benefits of 3D-CNN over 2D-CNN were limited. These findings highlight how leveraging the extra information captured in remote sensing data that includes more spectral information and multiple timepoints can depend on use of more complex analytical approaches. Successful application of this approach for lodging severity screening in biomass sorghum provides proof-of-concept for a tool to rapidly assess a trait that is of significant interest to breeders because it impacts harvestable yield, but is very challenging to assess at scale by traditional methods of visual inspection.

In recent years, agricultural studies have increasingly made use of a variety of imaging methods based on UAV-based remote sensing data. In the domains of plant breeding [33] and precision agriculture [34], this growing popularity reflects the need for accurate assessment of crop traits: (1) across many genotypes and at spatial scales that are too large for manual measurement to be efficient, yet (2) at spatial scales where the resolution and number of features (spectral and/or geometric) of imagery from piloted aircraft or satellites can be inadequate. Independent analysis of data from individual dates of image collection from UAVs has been the most common approach to date. For example, quantitative thresholding approaches have been applied in this way to aerial images to assess lodging severity in barley ($R^2 = 0.94$) [35] and lodging rate in maize ($R^2 = 0.50$) [36]. Success has also been achieved in estimating lodging severity by training and applying CNN analysis to a combined set of images from two developmental stages of wheat [20]. These studies demonstrate the potential for rapid assessment of lodging by UAV, but are limited by the need for manual intervention to select optimal threshold values in the workflow and by not leveraging the temporal variation in analysis of the data. They may also provide an early indication that estimating lodging severity is more challenging in tall, lower density, and coarsely-textured canopies such as maize relative to shorter, more uniformly dense and fine textured canopies such as rice, wheat and barley. If so, biomass sorghum is a good subject for testing methods that can cope with a difficult use case.

The results of the 2D-CNN analysis of images from different individual dates during the growing season provided valuable context for interpretation of the 3D-CNN results. Lodging severity could be predicted with reasonable accuracy when using images collected one day after the lodging event (Figure 5a). Unsurprisingly, there was almost no useful information to be gained from 2D-CNN analysis of images collected prior to the lodging event. And, the predictive ability of 2D-CNN analysis declined significantly and progressively when applied to images collected six and then fifteen days after the lodging event.

The superior performance of a 3D-CNN over a 2D-CNN for estimating lodging severity appears to depend on leveraging a “before and after” comparison of the event i.e. prediction accuracy measured in terms of R^2 , MAE or RMSE was best when data from 1 or 2 dates before and after the lodging event were analyzed (Figure 5). This is logical given that a change in the CSM (i.e. height profile) or spectral reflectance of the canopy before and after the lodging event would provide valuable additional information beyond the variation in geometric and spectral signals across plots on a single date once lodging had occurred. This interpretation is supported by 3D-CNN analysis of a time-course of images

from three dates after the lodging event failing to outperform the best 2D-CNN analysis (Figure 5).

Nevertheless, as long as the “before and after” comparison was made, the performance of the 3D-CNN was relatively insensitive to the number of sampling dates used prior to the lodging event (Figure 5). Even analysis of data from nine dates between 31 to 83 DAP allowed lodging severity prediction almost as accurate ($MAE = 9.99\%$, $RMSE = 12.68\%$, $R^2 = 0.72$) as when only the two dates before and after the lodging event were analyzed ($MAE = 9.65\%$, $RMSE = 12.40\%$, $R^2 = 0.75$). This suggests that significant error is not introduced into the 3D-CNN analysis when data from early in the season that has low information value with respect to lodging is included. And, as a practical consideration, it allows effective analysis of lodging without manual intervention to select dates of data collection nearest to the lodging event.

The notion that 3D-CNN analysis would improve UAV-based phenotyping beyond what 2D-CNN is capable of is based, in part, on recent satellite remote sensing research. Satellites are an ideal source of automatically generated time courses of remote sensing data [37]. And, as an example, inclusion of a third convolution layer in CNN architecture to integrate spatial and time dimensions in satellite data enabled improved crop classification at a regional scale [26]. 3D-CNN analysis of spatio-temporal models has been applied to UAV remote sensing for only a limited number of traits and crops. The applicability of 3D-CNN analysis of spatio-temporal data to predict yield [16] is consistent with yield being a product of compound growth and dynamic interactions with the environment over the entire growing season. But, the black-box nature of CNN analysis makes it hard to pinpoint which elements of the 3D-CNN and time-course of imagery drives enhancements in performance. By contrast, the abrupt nature of lodging allows the current study to provide a case study of how 3D-CNN analysis can enhance power through inclusion of only two timepoints of data “before and after” an event of interest in the life of the crop.

This study reveals the value of combining spectral and geometric information in feature-rich image data that are necessary to better leverage the benefits of 3D-CNN over 2D-CNN for estimation of lodging severity (Figure 5). Spectral information has previously proved to be a relevant descriptor of lodging in crops such maize [35], wheat [23], and barley [9]. The visual and near infrared regions of the spectrum have been the most common proxies for lodging detection [22] and lodging severity [21]. Geometric features derived from the photogrammetric reconstruction of the canopy have been less commonly used in these tasks [4] [2] [38]. But, the “before and after” comparison of geometric information in CSM was crucial for the enhanced predictive ability of 3D-CNN over 2D-CNN. This parallels and extends the successful integration of geometric and spectral information to improve predictions of traits such as biomass [12] [39] and plant height [40].

It was also notable that the benefit of integrating geometric and spectral features was greater for lodging severity than for lodging detection [36], regardless of the CNN architecture considered. This is consistent with previous findings [41,42], where the spectral information alone proved highly efficient in detecting lodging as pixel-wise classification [20]. And, this is probably, in part, explained by the fact that the geometric features are highly sensitive and less prone to signal saturation than spectral features [43], which becomes relevant when considering small differences in the degree of lodging severity between plots under dense canopy conditions.

According to the activation maps, both CNN architectures utilized similar regions of images to determine lodging detection and severity. But, the greater predictive power of the 3D-CNN did coincide with larger and more refined activation regions in the images compared to the 2D- models maps. In this regard, it can be argued that best 3D- model is able to extract additional informative descriptors from the spatial-temporal features during the 3D- convolution operations and, thereby, produce an improved prediction.

Even though we utilized manual ground-truth to build the models, we recognize that visual observation and quantification by walking through the alleys and plots after lodging was subject to errors that may affect the performance of the model. These could be improved by utilizing ground based imaging methods such as LIDAR for ground-truthing, but this was beyond the scope of our study. Future work might further exploit additional ground truth data of a quantitative nature for the use of spatio-temporal CNN architectures to better understand the capacity of certain genotypes to recover after lodging damage. Additional analyses are also needed to determine the impact of the spatial resolution of imagery and how to reduce processing time of data preprocessing and modelling implementation of 3D-CNN architectures.

5. Conclusions

This study aimed to explore the potential improvements in crop phenotyping with UAV remote sensing imaging that could be made with 3D-CNN versus 2D-CNN. Lodging detection and lodging severity in biomass sorghum were chosen as target traits for the case study because their short-term, stochastic nature present a distinct set of challenges and opportunities for interpretation versus traits such as yield which are determined over an entire growing season. In addition, lodging in highly productive bioenergy feedstock grasses and other cereal crops is agronomically important and inefficient to assess by traditional methods. It has previously been shown that lodging detection prediction with 2D-CNN is feasible and produces results accurate enough to monitor lodging. This study shows that adding explicit analysis of time as an additional dimension of variation in the spectral and geometric features imagery derived from UAV flights produces: (1) a modest improvement in the predictions of lodging detection, but (2) significant gain on the more challenging task of estimating quantitative variation in lodging severity. Using a small number of data collection dates before and after the lodging event provides enough data to outperform single time point model prediction in both tasks. But, using images from many dates across the growing season does not impair predictive ability substantially. This work partly addresses the need for advanced modelling techniques that can take advantage of recently improved spatial resolution, ease of collecting data, and cloud storage and processing for UAV remote sensing.

Supplementary Materials: The following are available online at <https://mdpi>

Author Contributions: S.V. and A.D.B.L. conceived the study, interpreted the data, and wrote the manuscript. S.V. collected, processed and analyzed data.

Funding: This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the U.S. Department of Energy.

Data Availability Statement: The datasets used and analysed during the current study are available from the corresponding author via Illinois Databank at https://doi.org/10.13012/B2IDB-0353090_V1

Acknowledgments: We thank Jeremy Ruther and Tim Mies at the Energy Farm at University of Illinois for technical assistance. We thank Patrick J. Brown, Taylor L. Pederson and Carl J. Bernacchi for providing germplasm and field establishment.

Conflicts of Interest: The authors declare no conflict of interest

References

1. Rocateli, A.C.; Raper, R.L.; Balkcom, K.S.; Arriaga, F.J.; Bransby, D.I. Biomass Sorghum Production and Components under Different Irrigation/Tillage Systems for the Southeastern U.S. *Ind. Crops Prod.* 2012, 36, 589–598, doi:10.1016/j.indcrop.2011.11.007.

2. Murakami, T.; Yui, M.; Amaha, K. Canopy Height Measurement by Photogrammetric Analysis of Aerial Images: Application to Buckwheat (*Fagopyrum Esculentum* Moench) Lodging Evaluation. *Comput. Electron. Agric.* 2012, 89, 70–75, doi:10.1016/j.compag.2012.08.003. 583–586
3. Li, X.; Li, X.; Liu, W.; Wei, B.; Xu, X. A UAV-Based Framework for Crop Lodging Assessment. *Eur. J. Agron.* 2021, 123, 126201, doi:10.1016/j.eja.2020.126201. 587–588
4. Chauhan, S.; Darvishzadeh, R.; Boschetti, M.; Pepe, M.; Nelson, A. Remote Sensing-Based Crop Lodging Assessment: Current Status and Perspectives. *ISPRS J. Photogramm. Remote Sens.* 2019, 151, 124–140, doi:10.1016/j.isprsjprs.2019.03.005. 589–591
5. Gomez, F.E.; Muliana, A.H.; Niklas, K.J.; Rooney, W.L. Identifying Morphological and Mechanical Traits Associated with Stem Lodging in Bioenergy Sorghum (*Sorghum Bicolor*). *BioEnergy Res.* 2017, 10, 635–647, doi:10.1007/s12155-017-9826-7. 592–594
6. Mizuno, H.; Kasuga, S.; Kawahigashi, H. Root Lodging Is a Physical Stress That Changes Gene Expression from Sucrose Accumulation to Degradation in Sorghum. *BMC Plant Biol.* 2018, 18, 2, doi:10.1186/s12870-017-1218-9. 595–597
7. Guo, J.R.; Fan, H.; Wang, B.S. Lodging Markedly Reduced the Biomass of Sweet Sorghum via Decreasing Photosynthesis in Saline-Alkali Field. *E3S Web Conf.* 2018, 38, 02016, doi:10.1051/e3sconf/20183802016. 598–600
8. Setter, T.L.; Laureles, E.V.; Mazaredo, A.M. Lodging Reduces Yield of Rice by Self-Shading and Reductions in Canopy Photosynthesis. *Field Crops Res.* 1997, 49, 95–106, doi:10.1016/S0378-4290(96)01058-1. 601–603
9. Vlachopoulos, O.; Leblon, B.; Wang, J.; Haddadi, A.; LaRocque, A.; Patterson, G. MAPPING BARLEY LODGING WITH UAS MULTISPECTRAL IMAGERY AND MACHINE LEARNING. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 2021, XLIII-B1-2021, 203–208, doi:10.5194/isprs-archives-XLIII-B1-2021-203-2021. 604–607
10. Fischer, R.A.; Stapper, M. Lodging Effects on High-Yielding Crops of Irrigated Semidwarf Wheat. *Field Crops Res.* 1987, 17, 245–258, doi:10.1016/0378-4290(87)90038-4. 608–609
11. Fernandes, S.B.; Dias, K.O.G.; Ferreira, D.F.; Brown, P.J. Efficiency of Multi-Trait, Indirect, and Trait-Assisted Genomic Selection for Improvement of Biomass Sorghum. *Theor. Appl. Genet.* 2018, 131, 747–755, doi:10.1007/s00122-017-3033-y. 610–612
12. Varela, S.; Pederson, T.; Bernacchi, C.J.; Leakey, A.D.B. Understanding Growth Dynamics and Yield Prediction of Sorghum Using High Temporal Resolution UAV Imagery Time Series and Machine Learning. *Remote Sens.* 2021, 13, 1763, doi:10.3390/rs13091763. 613–615
13. Habyarimana, E.; Piccard, I.; Catellani, M.; De Franceschi, P.; Dall'Agata, M. Towards Predictive Modeling of Sorghum Biomass Yields Using Fraction of Absorbed Photosynthetically Active Radiation Derived from Sentinel-2 Satellite Imagery and Supervised Machine Learning Techniques. *Agronomy* 2019, 9, 203, doi:10.3390/agronomy9040203. 616–619
14. Han, D.; Yang, H.; Yang, G.; Qiu, C. Monitoring Model of Corn Lodging Based on Sentinel-1 Radar Image. In *Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA)*; November 2017; pp. 1–5. 620–622
15. Peake, A.S.; Bell, K.L.; Carberry, P.S.; Poole, N.; Raine, S.R. Vegetative Nitrogen Stress Decreases Lodging Risk and Increases Yield of Irrigated Spring Wheat in the Subtropics. *Crop Pasture Sci.* 2016, 67, 907, doi:10.1071/CP16052. 623–625
16. Acorsi, M.G.; Martello, M.; Angnes, G. IDENTIFICATION OF MAIZE LODGING: A CASE STUDY USING A REMOTELY PILOTED AIRCRAFT SYSTEM. *Eng. Agric.* 2019, 39, 66–73, doi:10.1590/1809-4430-eng.agric.v39nep66-73/2019. 626–628

17. Nevavuori, P.; Narra, N.; Linna, P.; Lipping, T. Crop Yield Prediction Using Multitemporal UAV Data and Spatio-Temporal Deep Learning Models. *Remote Sens.* 2020, 12, 4000, doi:10.3390/rs12234000. 629–631
18. Hosoi, F.; Nakai, Y.; Omasa, K. Estimating the Leaf Inclination Angle Distribution of the Wheat Canopy Using a Portable Scanning Lidar. *J. Agric. Meteorol.* 2009, 65, 297–302, doi:10.2480/agrmet.65.3.6. 632–634
19. Liu, H.Y.; Yang, G.J.; Zhu, H.C. The Extraction of Wheat Lodging Area in UAV's Image Used Spectral and Texture Features. *Appl. Mech. Mater.* 2014, 651–653, 2390–2393, doi:10.4028/www.scientific.net/AMM.651-653.2390. 635–637
20. Zhao, B.; Li, J.; Baenziger, P.S.; Belamkar, V.; Ge, Y.; Zhang, J.; Shi, Y. Automatic Wheat Lodging Detection and Mapping in Aerial Imagery to Support High-Throughput Phenotyping and In-Season Crop Management. *Agronomy* 2020, 10, 1762, doi:10.3390/agronomy10111762. 638–641
21. Sun, Q.; Sun, L.; Shu, M.; Gu, X.; Yang, G.; Zhou, L. Monitoring Maize Lodging Grades via Unmanned Aerial Vehicle Multispectral Image. *Plant Phenomics* 2019, 2019, doi:10.34133/2019/5704154. 642–644
22. Zhang, Z.; Flores, P.; Igathinathane, C.; L. Naik, D.; Kiran, R.; Ransom, J.K. Wheat Lodging Detection from UAS Imagery Using Machine Learning Algorithms. *Remote Sens.* 2020, 12, 1838, doi:10.3390/rs12111838. 645–647
23. Koh, J.C.O.; Spangenberg, G.; Kant, S. Automated Machine Learning for High-Throughput Image-Based Plant Phenotyping. *Remote Sens.* 2021, 13, 858, doi:10.3390/rs13050858. 648–650
24. Jiang, Y.; Li, C. Convolutional Neural Networks for Image-Based High-Throughput Plant Phenotyping: A Review. *Plant Phenomics* 2020, 2020, doi:10.34133/2020/4152816. 651–652
25. Yang, M.-D.; Tseng, H.-H.; Hsu, Y.-C.; Tsai, H.P. Semantic Segmentation Using Deep Learning with Vegetation Indices for Rice Lodging Identification in Multi-Date UAV Visible Images. *Remote Sens.* 2020, 12, 633, doi:10.3390/rs12040633. 653–655
26. Ji, S.; Zhang, C.; Xu, A.; Shi, Y.; Duan, Y. 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. *Remote Sens.* 2018, 10, 75, doi:10.3390/rs10010075. 656–658
27. Li, Y.; Zhang, H.; Shen, Q. Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network. *Remote Sens.* 2017, 9, 67, doi:10.3390/rs9010067. 659–660
28. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional Neural Networks: An Overview and Application in Radiology. *Insights Imaging* 2018, 9, 611–629, doi:10.1007/s13244-018-0639-9. 661–663
29. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning Spatiotemporal Features with 3D Convolutional Networks. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*; IEEE: Santiago, Chile, December 2015; pp. 4489–4497. 664–667
30. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization. *ArXiv151204150 Cs* 2015. 668–669
31. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In *Proceedings of the Computer Vision – ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer International Publishing: Cham, 2014; pp. 818–833. 670–672
32. Team, K. Keras Documentation: Grad-CAM Class Activation Visualization Available online: https://keras.io/examples/vision/grad_cam/ (accessed on 1 December 2021). 673–674

33. Watt, M.; Fiorani, F.; Usadel, B.; Rascher, U.; Muller, O.; Schurr, U. Phenotyping: New Windows into the Plant for Breeders. *Annu. Rev. Plant Biol.* 2020, 71, 689–712, doi:10.1146/annurev-arplant-042916-041124.
34. Messina, G.; Modica, G. Applications of UAV Thermal Imagery in Precision Agriculture: State of the Art and Future Research Outlook. *Remote Sens.* 2020, 12, 1491, doi:10.3390/rs12091491.
35. Chu, T.; Starek, M.J.; Brewer, M.J.; Murray, S.C.; Pruter, L.S. Assessing Lodging Severity over an Experimental Maize (*Zea Mays* L.) Field Using UAS Images. *Remote Sens.* 2017, 9, 923, doi:10.3390/rs9090923.
36. Wilke, N.; Siegmann, B.; Klingbeil, L.; Burkart, A.; Kraska, T.; Muller, O.; van Doorn, A.; Heinemann, S.; Rascher, U. Quantifying Lodging Percentage and Lodging Severity Using a UAV-Based Canopy Height Model Combined with an Objective Threshold Approach. *Remote Sens.* 2019, 11, 515, doi:10.3390/rs11050515.
37. Liu, S.; Marinelli, D.; Bruzzone, L.; Bovolo, F. A Review of Change Detection in Multitemporal Hyperspectral Images: Current Techniques, Applications, and Challenges. *IEEE Geosci. Remote Sens. Mag.* 2019, 7, 140–158, doi:10.1109/MGRS.2019.2898520.
38. Yang, M.-D.; Huang, K.-S.; Kuo, Y.-H.; Tsai, H.P.; Lin, L.-M. Spatial and Spectral Hybrid Image Classification for Rice Lodging Assessment through UAV Imagery. *Remote Sens.* 2017, 9, 583, doi:10.3390/rs9060583.
39. Jiang, Q.; Fang, S.; Peng, Y.; Gong, Y.; Zhu, R.; Wu, X.; Ma, Y.; Duan, B.; Liu, J. UAV-Based Biomass Estimation for Rice-Combining Spectral, TIN-Based Structural and Meteorological Features. *Remote Sens.* 2019, 11, 890, doi:10.3390/rs11070890.
40. Bendig, J.; Yu, K.; Aasen, H.; Bolten, A.; Bennertz, S.; Broscheit, J.; Gnyp, M.L.; Bareth, G. Combining UAV-Based Plant Height from Crop Surface Models, Visible, and near Infrared Vegetation Indices for Biomass Monitoring in Barley. *Int. J. Appl. Earth Obs. Geoinformation* 2015, 39, 79–87, doi:10.1016/j.jag.2015.02.012.
41. Du, M.; Noguchi, N. Monitoring of Wheat Growth Status and Mapping of Wheat Yield's within-Field Spatial Variations Using Color Images Acquired from UAV-Camera System. *Remote Sens.* 2017, 9, 289, doi:10.3390/rs9030289.
42. Constantinescu, C.A.; Herbei, M.V.; Manea, D.; Sala, F. ANALYSIS OF SOME DEFICIENCIES IN CROPS OF WHEAT AND BARLEY BASED ON TERRESTRIAL AND AERIAL IMAGES. 2017, 10.
43. Gitelson, A.A. Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *J. Plant Physiol.* 2004, 161, 165–173, doi:10.1078/0176-1617-01176.