TECHNIQUES AND METHODS FOR IMPROVED EFFECTIVENESS AND ACCURACY IN COMPUTER VISION APPLICATIONS

A Thesis

by

KRISHNA CHAITANYA GADEPALLY

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Kevin Nowka
Committee Members,	Nick Duffield
	Dezhen Song
Head of Department,	Miroslav Begovic

May 2023

Major Subject: Electrical Engineering

ABSTRACT

Computer vision and image processing algorithms work well under strong conditions. Unexpected computer vision results may harm pipeline modules, reducing solution efficacy. A predictor framework that was simultaneously trained with a hardness predictor network to mitigate such effects was used. The performance of this framework is guaranteed to be better than that of images with lower "hardness" levels, which improves the efficiency of modules using computer vision algorithms. There is a clear understanding of why neural networks perform well. Images with less complex patterns produced relatively stronger activations in the initial convolutional layers. This study also examined how well a convolutional neural network worked when training, validating, and testing eliminated irrelevant convolutional layers for a single image. This enhanced CNNs' efficiency.

Predictive modeling of complex time-dependent processes is erroneous in many image-based machine learning applications due to the absence of early data. Predictive agriculture uses this frequently. From early growth until harvest, cotton crop factors must be monitored. Because cotton crop output is significantly correlated with growth parameter management during a cultivation season, researchers are interested in developing forecasting models to predict canopy and vegetative indices. This study utilized a multi-layer stacked LSTM model on cotton plant canopy and vegetative attributes obtained from unmanned aerial vehicle survey images from the 2020 growing season. Based on the 2021 cultivation year canopy and vegetative index data, the

weights of the final few layers of the trained LSTM model were fine-tuned to predict canopy and vegetative indices. Deep Transfer Learning improved UAS canopy and vegetative indices’ accuracy.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professor Kevin Nowka and Professor Nick Duffield of the Department of Electrical and Computer Engineering and Professor Dezhen Song of the Department of Computer Science and Engineering.

The data analyzed for 3.3 was provided by Texas A&M AgriLife.

All other work conducted for the thesis was completed by the student independently.

Funding Sources

Graduate study was supported by a teaching assistantship from Texas A&M University.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION*

This study deals with techniques and methods for improved effectiveness and increased accuracies in computer vision applications. Algorithms for computer vision and image processing perform well when certain conditions are met. Not all types of inputs are anticipated to perform well for computer vision algorithms. For instance, computer vision algorithms may not produce the best results for images that are overly noisy. Solution effectiveness may suffer due to unexpected results from the computer vision module having detrimental effects on subsequent pipeline modules. A predictor framework that was simultaneously trained with a hardness predictor network to mitigate such effects was developed. The hardness predictor network outputs hardness values for inputs. The performance of this framework is guaranteed to be better than that of images with lower "hardness" levels, which improves the efficiency of modules using computer vision algorithms.

Likewise, for a long time, neural networks were viewed as mysterious black boxes. To aid in matching of model to data, Zeiler et al. [1] identified the characteristics of a picture that were crucial for convolutional layers at various points in the network. Deconvolutional networks were used for this. This work looked at how well a

---

* Parts of this chapter are reprinted with permission from "A Deep Transfer Learning based approach for forecasting spatio-temporal features to maximize yield in cotton crops" by K. C. Gadepally, S. B. Dhal, M. Bhandari, J. Landivar, S. Kalafatis and K. Nowka, 2023 57th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 2023, pp. 1-4, doi: 10.1109/CISS56502.2023.10089748.

convolutional neural network performed when the training, validating, and testing processes omitted convolutional layers that were comparatively irrelevant for a specific image (i.e., the image did not create one of the strongest activations). This method improved the performance of CNNs.

Finally, in many applications of image-based machine learning, predictive modeling of complex time-dependent processes suffer in predictive accuracy due to lack of data early in the time series. This is especially prevalent in predictive agriculture use cases. In this work, we concentrated on developing a predictive yield modeling machine learning methodology for cotton production. In the United States, cotton is a significant crop for economic reasons. It is crucial to monitor cotton crop growth parameters throughout the growing season, from early season growth until harvest. Given that management decisions made to control growth parameters during a cultivation season have a significant impact on cotton crop production, researchers are interested in applying forecasting models to predict future values of the canopy and surplus green index. The earlier yield can be predicted accurately, the better. This study applied a multi-layer stacked LSTM model using data on cotton plant canopy and vegetative features and vegetative indices, including canopy cover, canopy height, and excess green index, which were obtained from unmanned aerial vehicle survey images from the 2020 growing season. In order to forecast the canopy and vegetative features from the 28th day of culture until the conclusion of the harvesting season, the weights of the final few layers of the trained LSTM model were fine-tuned based on the 2021 and 2022 cultivation years' canopy and vegetative index data. This was done using a Deep

Transfer Learning technique [2], which boosted the accuracy of canopy and vegetative features obtained from UAS data.

2. BACKGROUND*

2.1. Realistic Predictors For Regression And Semantic Segmentation

Systems for computer vision (CV) are used in fields where error margins are getting smaller. One such sector is the military. Unmanned aircraft systems (UAS), for instance, are utilized in this industry for reconnaissance, attacks, and surveys. A UAS has navigation, safety, control, and perception subsystems. Perception systems frequently include CV systems.

The framework and design of CV algorithms are based on strict presumptions. Decent image quality, minimal noise, good vision, and little motion blurring are only a few examples of these assumptions regarding the caliber of input photographs. A UAS's CV system should have undergone thorough validation and testing using both "real world" data and datasets that are considered industry standards. It is a fact that the "real world" data utilized to test computer vision systems only represents a small portion of the vast array of sensory inputs that are available to UAS while they are in use. Such a diverse range of sensory inputs is beyond the capabilities of CV systems. Other subsystems receive the output of perception subsystems. As a result, it is possible to say that perception systems interact with the environment and other subsystems.

---

* Parts of this chapter are reprinted with permission from "A Deep Transfer Learning based approach for forecasting spatio-temporal features to maximize yield in cotton crops" by K. C. Gadepally, S. B. Dhal, M. Bhandari, J. Landivar, S. Kalafatis and K. Nowka, 2023 57th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 2023, pp. 1-4, doi: 10.1109/CISS56502.2023.10089748.

Some of the novel and uncharted data points might cause CV systems to react in ways that have disastrous consequences for other subsystems in the pipeline. Despite built-in safety features, this can still occur. This motivates us to employ this technique for feature extraction from the imagery dataset, which aids in the development of a more reliable model for regression on images as well as for semantic segmentation while enhancing the bias-variance tradeoff. Identifying each pixel in an image from a preset set of classes is the problem of semantic image segmentation.

A number of studies, including Zhang et al. [3] and Daftry et al. [4], developed frameworks to assess the CV module's suitability for making decisions on the current set of imagery inputs. Both of these approaches rely on post-hoc evaluations of the predictor performance, where a regressor or classifier is simply taught from its errors. The framework presented by Wang et al. [5] is different from the two presented above. It is independent of post-hoc evaluations of the predictor's effectiveness. The unique feature of this framework is the simultaneous training of two networks. The goal of our method is to simultaneously train a predictor framework for image segmentation or regression and a hardness predictor network that generates hardness values for inputs.

There are two networks trained simultaneously. A Hardness Predictor network, which outputs hardness values for inputs, is trained simultaneously with the predictor our approach. The hardness predictor is an independent network that is trained in tune with the predictor but operates independently after training. The roles of hardness predictor network and predictor cannot be performed by a single network. It cannot be trained independently of the predictors. Attempts to train hardness predictor network and predictor independently of each other or non-simultaneously would yield suboptimal outcomes. A single network cannot serve as both a predictor and a hardness predictor network. Because hardness values are arbitrary, attempts to produce training data for hardness predictors would be unsuccessful. Therefore, attempting to train the hardness predictor network and predictor separately or not at the same time would result in less-than-ideal results. This work extended the machine learning classification framework of Wang et al. to develop a methodology for specifying realistic predictors for additional machine learning frameworks for semantic segmentation and regression.

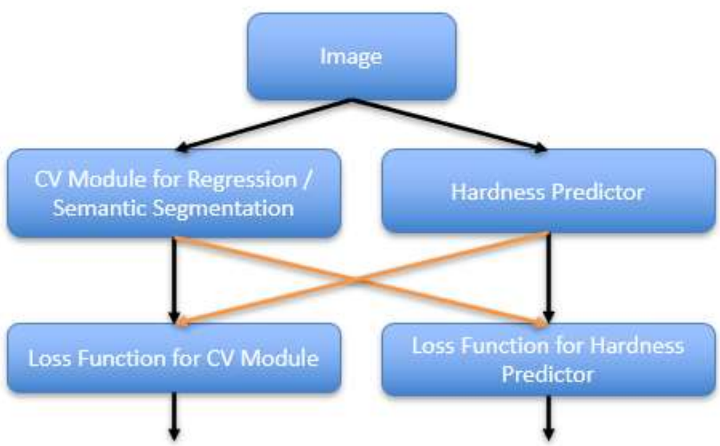

Figure 1 Block Diagram For Hardness Prediction (Regression And Semantic Segmentation)

2.2. Systematic Model Complexity Reduction by Elimination of Irrelevant Layers

Zeiler et al. [1] have provided a good explanation of why neural networks operate successfully. Zeiler et al. [1] demonstrated using deconvolutional networks that images with simpler patterns generated substantially stronger activations in the initial convolutional layers. In Zeiler et al. [6], the mathematics underlying deconvolutional networks were thoroughly illustrated.

On the other hand, deeper convolutional layers showed considerably stronger activations in images with more intricate patterns and forms.

This implied that for simple images, the earliest convolutional layers were significantly more significant, whereas for complex images, the inverse was true.

Convolutional neural networks may perform better as a result of this information. For instance, if the network struggled with images of comparatively lower complexity, the settings relating to the early convolutional layers may be changed appropriately. These modifications might alter the number of feature maps, kernel size, or activation function.

This study investigated the performance of a convolutional neural network when the training, validating, and testing phases are skipped for convolutional layers that are comparatively irrelevant for a given image (i.e., the image does not yield one of the strongest activations).

2.3. Deep Transfer Learning For Forecasting Spatio-Temporal Features For Yield Prediction In Cotton Crops

One of the most significant agricultural products and cash crops in the United States and around the globe is cotton. After China and India, the United States is the third-largest cotton producer in the world. In 2021, cotton was produced in the United

States in 17.62 million bales. By 2022, 12.2 million acres of cotton are anticipated to be farmed in the United States, with around 7 million of those acres being planted in Texas alone [7]. Cotton plants can be examined and used as a test subject for managerial, environmental, and genetic factors. Sensors are used to measure the effects of such interaction on plant growth and development. In the past, Zhao et al. [8] forecasted growth and production in irrigated cotton using canopy spectral reflectance and leaf area index, highlighting the significance of reflectance data from early blooming in yield prediction. According to this pattern, unmanned aerial systems may easily obtain and evaluate these reflectance characteristics and agronomic parameters (UAS).

In order to plan and manage production wisely in sustainable agriculture, it is essential to have reliable and accurate forecasting models. Historically, a wide range of models have been built using historical information amassed over time. To build agricultural pest and disease warning systems, various models have been developed and evaluated, including within-year models, between-year models, spectral data models, and farmer assessment models [9]. To estimate yield in non-linear olive orchards in mountainous terrains, Normalized Difference Vegetation Indices (NDVI) were recovered from orthomosaics using feature extraction techniques [10].

Recently, some work has been done to predict the yield using time-series data. As inputs to the linear models that predicted final soybean output, the data was modeled to include the sum of the NDVI and EVI vegetation indices for the entire growth period, up to 70 days of growth phase, up to 85 days of growth phase, and up to 98 days of growth phase [11]. Similar to this, time-series data was obtained to forecast sugarcane harvest at the field scale by combining information from crop growth models, expert knowledge, and satellite images [12].

Researchers at Texas A&M Agrilife Center compared various Long Short-Term Memory (LSTM) models to predict canopy cover, canopy height, and excess green indices of cotton crops using training datasets produced by K-Means clustering and Dynamic Time Warping (DTW) techniques of data modeling. The predictions were of limited accuracy.

With the help of a multi-layer stacked LSTM model that was trained using the canopy and vegetative indices collected for the cultivation year 2020, we employed a deep transfer learning strategy in this study. The weights of the final few layers of the stacked LSTM were adjusted based on the data gathered throughout the 2021 and 2022 cultivation years by freezing the weights of the first few layers. From the 28th day of cultivation through the conclusion of the cultivation period, the canopy and vegetative indices were predicted, and based on these values, in-season management choices were made in order to maximize the yield at the end of the cultivation period.

3. METHODS AND RESULTS*

3.1 Realistic Predictors For Regression And Semantic Segmentation

3.1.1. Realistic Predictors For Regression

In this framework, the convolutional networks of AlexNet [13] and LeNet [14] with 2 and 4 dense layers were used for regression and hardness prediction respectively. Both the networks were fully trained. Multiple splits of the Places database [15] was used for training, validation, and testing the results. Weighted Mean Squared Error (W-MSE) was used as the loss function for the regressor which is computed as follows in Eqn (1).

$$WMSE_{AlexNet} = \sum_i o^i_{hardness}\left(o^i_{regressor,predicted} - o^i_{regressor,ideal}\right)^2 \qquad (1)$$

Outputs of the hardness predictor network were the weights. Binary Cross Entropy (BCE) of the hardness predictor network was used as the error metric and was compared against the ideal outputs which are calculated by taking the sigmoid of the regressor outputs. The calculated BCE was then used as the loss function for the hardness predictor which is computed as follows in Eqn (2).

---

* Parts of this chapter are reprinted with permission from "A Deep Transfer Learning based approach for forecasting spatio-temporal features to maximize yield in cotton crops" by K. C. Gadepally, S. B. Dhal, M. Bhandari, J. Landivar, S. Kalafatis and K. Nowka, 2023 57th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 2023, pp. 1-4, doi: 10.1109/CISS56502.2023.10089748.

$$BCE_{LeNet} = -\sum_i sigmoid(o^i_{regressor,predicted}) * \log(o^i_{hardness}) \quad (2)$$

3.1.2. Realistic Predictors For Semantic Segmentation

In this framework, the convolutional layers of SegNet [16] and LeNet with 0 and 4 dense layers were used for semantic segmentation and hardness prediction respectively. The hardness prediction framework used in this case followed the same framework as the one used for extracting the realistic predictors for regression. Several splits of the Oxford-IIIT Pet Database [17] and PASCAL VOC-2012 Database [18] were used for training the dual architecture.

The output of SegNet was an image of the same height and width as the input but with three channels. Softmax activation function was applied to the output. For each pixel in the output, the channel number (0, 1 or 2 in this case) with the highest probability value was the output value for that pixel. The performance metric for SegNet was the probability that for a pixel, the channel number at which the high probability occurs is equal to the value of the pixel at the same position in the ground truth. Weighted Sparse Categorical Cross Entropy (WSCCE) with outputs of the Hardness Predictor Network as weights was used as the loss function for SegNet. Those images with high Hardness values were incentivized by the loss function to improve their performance. The WSCCE of the SegNet framework is calculated as follows in Eqn (3).

$$WSSCE_{SegNet} = -(o^i_{hardness}) \quad * \sum_i (o^i_{segmentation}) * \log(o^i_{segmentation,predicted}) \qquad (3)$$

On the other hand, similar to that of the regression framework, BCE was used as the loss function for the Hardness predictor. For each image, the performance metric subtracted from 1 was used as the ground truth value, which was a good proxy indicator. Greater the performance metric (which lies between 0 and 1), lower is the value of hardness for that image. BCE for the hardness predictor network in the case of semantic segmentation is represented as in Eqn (4).

$$BCE_{LeNet} = -\sum_i (1 - o^i_{hardness}) * \log(o^i_{hardness}) \qquad (4)$$

For our experimentation, the Grace cluster belonging to Texas A&M University HPRC was used which was a 48 core Linux machine with 2 nodes to obtain the results. The Places database (comprising 3000 data points) was used for training, validation, and testing. A batch size of 256 was used for training both the networks. L1 and L2 regularizations were used for training the regressor. The dual network was trained for 100 epochs. After training, it was found that the hardness values for the test set ranged from 0.328 to 0.977.

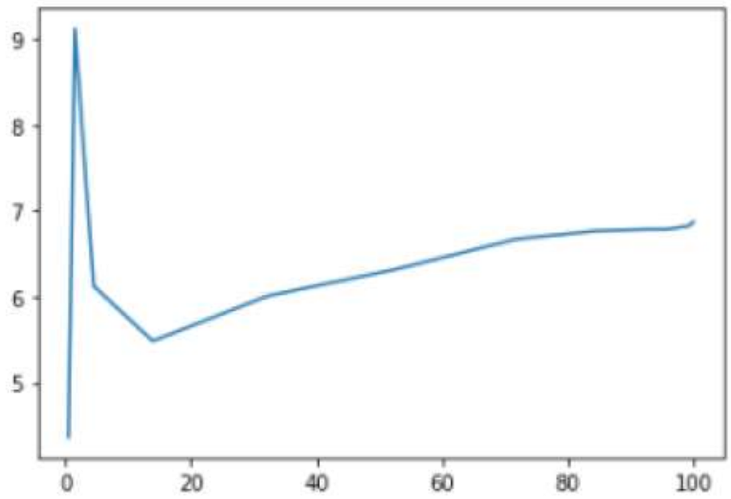

Figure 2 Mae Vs Percentage Of Data Points From The Test Set With The Lowest Hardness Values Used For Testing

The results are as shown in Figure 2. The x-axis represents the percentage of data points from the test set with the lowest hardness values used for testing. The y-axis represents the Mean Absolute Error (MAE) scores of those data points from the test set. For example, 60 on the x-axis means that data points with 40 percent of the highest hardness values were removed from testing. Only the bottom 60 percent were used.

Oxford-IIIT Pet Dataset was used for training, validation, and testing our analysis on semantic segmentation. The SegNet-LeNet dual network was trained for 100 epochs SegNet-LeNet dual network was trained for 100 epochs with a batch size of 32.

The validation set was given as input to the Hardness Predictor network after training. The corresponding hardness values ranged from 0.495 to 0.644. Images that were easier to segment had relatively lower hardness values (outputs of Hardness Predictor), while those that were relatively harder to segment had higher values.

Figure 3 shows how the custom performance metric varies with the cut-off value of Hardness scores (Oxford-IIIT Pets Dataset). We expect the custom performance metric value to decrease with increase in cut-off.

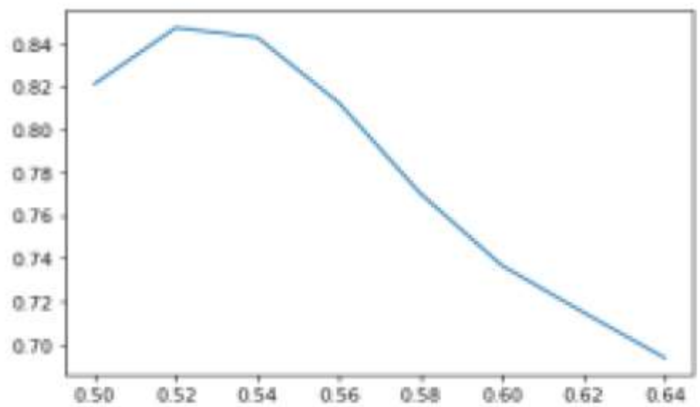

Figure 3 Custom Performance Metric Value Vs Hardness Values Cutoff For Oxford-IIIT Pets Dataset

Here, in case of semantic segmentation, two different cut-offs a and b are considered, where a > b. Those images with hardness values between a and b would be additionally included for testing in the case of the cut-off changing from b to a. Clearly, additional harder images are included when cut-offs increase, and the performance is expected to worsen.

Hardness value ranged from 0.3374 to 0.5182 in the case of PASCAL VOC-2012 Dataset.

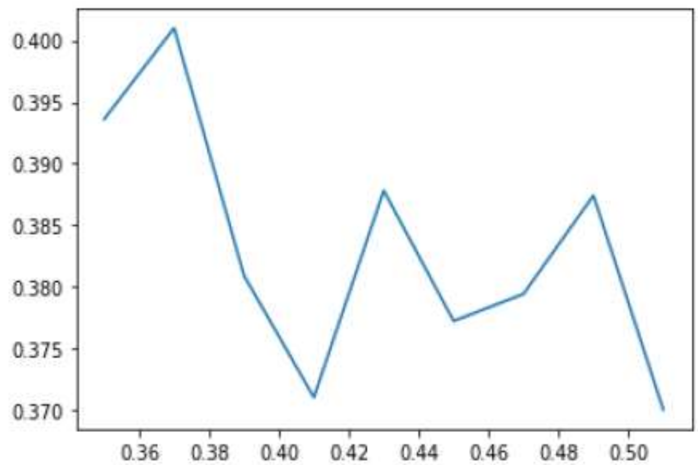

Figure 4 Custom Performance Metric Value Vs Hardness Values Cutoff For Pascal VOC-2012 Dataset

3.2 Systematic Model Complexity Reduction by Elimination of Irrelevant Layers

In order to train, validate, and test a three-layered convolutional neural network "CNN 1" with one dense layer, the MNIST dataset [19] was employed. It is significant to note that just the number of feature mappings distinguished the three convolutional layers; there were no max pooling layers. The first ten feature maps for each of the three convolutional layers were chosen. Images that provided the top 10 activations for each of the three convolutional layers for each of the ten feature maps were picked. This indicated that 100 photos were selected for each of the three convolutional layers.

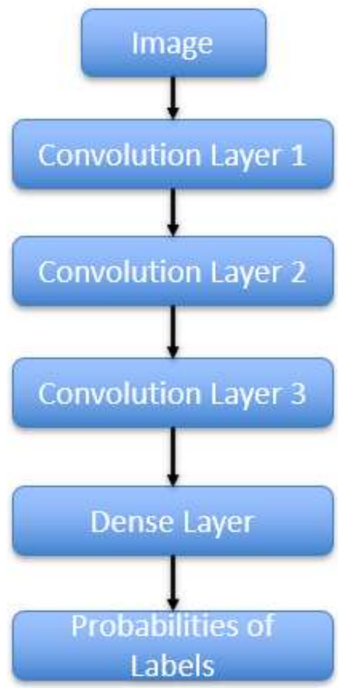

Figure 5 Architecture Of CNN 1 In Which All Convolutional Layers Are Utilized

Depending on which convolutional layer an image produced one of the top 10 activations, it was given a label. Since that convolutional layer was important in processing the image, the higher label was chosen in the event of repetition.

Based on these training data, a convolutional neural network "CNN 2" was trained to predict the convolutional layer at which the input image resulted in one of the strongest activations. CNN 2 had three convolutional layers and one dense layer. CNN 2 had 128, 64 and 32 neurons for its three convolutional layers respectively. ReLU activation function was used for all three activation layers.

The final phase was training a second convolutional neural network with the aforementioned images and labels “CNN 3”. CNN 1 and CNN 3 have the same architecture. It should be noted that the designations 0, 1, and 2 were all used three times. The images with the strongest activations in the first convolutional layer were given a score of 0. Similar to labels 1 and 2, labels 1 and 2 were for images that, in the second and third convolutional layers, respectively, produced the strongest activations.

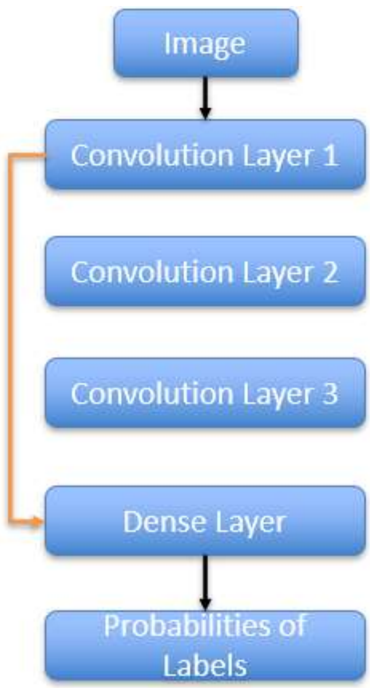

Figure 6 Architecture Of CNN 3 In Which Not All Convolutional Layers Are Utilized

Weights were initialized for the convolutional neural network (CNN 3). The second and third convolutional layers for pictures with label 0 were not included in the forward pass. Only the thick layer and the first convolutional layer were used in the forward pass. Thus, weights for the dense layer and the first convolutional layer were modified. The second and third convolutional layer's weights were not changed. The third convolutional layer's weights for the images corresponding to label 1 were similarly ignored, and other weights were changed following the forward pass that skipped the third convolutional layer. No layer was skipped in the forward pass for the photos with label 2, therefore all weights were updated.

It was necessary to ensure that all the convolutional layers in CNN 3 had the same height and width for this to occur, i.e., skipping convolutional layers in forward pass.

Using the RMSprop optimizer, ReLU activation function, sparse categorical cross entropy loss function, and accuracy performance metric for 20 iterations, the aforementioned convolutional neural network was trained. The CIFAR-10 dataset underwent the same replication procedure as above.

CNN 2 was trained based on this training data to predict the convolutional layer at which the input image produced one of the strongest activations had a test accuracy of 62 percent.

Those results were used to train, validate, and test the convolutional neural network mentioned in the previous section (CNN 3). It had a test accuracy of 62 percent. Those results were used to train, validate, and test the convolutional neural network mentioned in the previous section (CNN 3). It had a test accuracy equal to 99.2 percent, marginally higher than 98 percent when a convolutional neural network was used to train, validate, and test the MNIST dataset without any caveats (CNN 1). This result was validated using the CIFAR-10 dataset too. There was a slight fall in performance by 1.1 percent when CNN 3 was used instead of CNN 1.

3.3 A Deep Transfer Learning Based Approach For Forecasting Spatio-Temporal Features To Maximize Yield In Cotton Crop

In this study, data on canopy variables, including as canopy cover, canopy height, and excess green index, that were acquired in the year 2020 were used to train a multi-layer stacked LSTM model. The weights of the last few layers of the trained LSTM model were adjusted based on the 2021 cultivation year canopy and vegetative index data to forecast the canopy and vegetative features from the fourth week of the harvesting season to the tenth week. This was accomplished utilizing a multi-layer stacked LSTM model and the Deep Transfer Learning approach.

3.3.1 Construction Of The Dataset

The experimental site is located at Driscoll in Corpus Christi, Texas, USA (latitude of 27°40'06.2''N and longitude of 97°97°41'22.6''W). The information was gathered at weekly intervals. Data was collected by taking raw photos while flying a UAS over a cotton field. The DJI Phantom 4 RTK (DJI-P4RTK) with an RGB sensor (20 M) was used in this experiment. The first flight took place shortly following the planting. This trip was critical because it was used to generate a digital terrain model (DTM) that was used to estimate canopy height from UAS imagery [20].

Following the completion of the UAS data collection, the imagery was photogrammetrically processed to provide dense 3D point cloud data. The structure from motion approach (SfM) was used to generate 3D point cloud data, which approximates a three-dimensional structure using two-dimensional images. A digital surface model (DSM) and orthorectify imagery were created using the 3D point cloud data. The final one was then utilized to make an orthomosaic.

There are numerous commercial and open-source SfM software options available for processing UAS imagery. The imaging datasets in this research were processed with the Agisoft Metashape PRO software version 1.8.2 (AgiSoft LLC, St. Petersburg, Russia).

In its workflow, the Agisoft Metashape program followed a sequence of phases (fig. 5). During the initial step of the structure from motion method, a low-level procedure termed feature detection was employed to extract features with different patterns from neighbor pictures. Camera pose (position and orientation of the camera in relation to a reference coordinate system) and scene geometry were recreated at the same time in this method. During the keypoint correspondence procedure, the detected characteristics were then matched across, allowing an estimate of the camera's calibration matrix and prediction of camera motion. Feature points were sometimes mismatched. When this occurred, keypoint filtering was useful in removing low-quality keypoint matches.

Georeferencing was also involved in the procedure. Ground control points (GCPs) were utilized to anchor the map to the ground, matching the position of the drone photos to the location of the measured GCP. To georeference the photos for this project, one GCP and real-time kinematics (RTK) were used. SfM bundle adjustment recreated the 3D environment using the matched/filtered keypoints and camera settings as part of the processing phase. The 3D point set was utilized to create a triangulated irregular network (TIN) and the DSM. In addition, the DSM was utilized to project every pixel in the photos in order to compute a geometrically corrected image mosaic with uniform scale [21].

After generating the DSM and orthomosaics for each After generating the DSM and orthomosaics for each flight, phenotyping data extraction commenced. A 10x10 meter grid was first created across the cotton field. The AgriLife center's Python programs were then used to do pixel analysis on the drawn boundaries. The plant height was then estimated by subtracting the pixel height-above-earth recorded during the first flight from the original ground layer. Finally, CHM was used to estimate the volume of objects on the ground by taking the height-above-ground value and multiplying it by the pixel area [22].

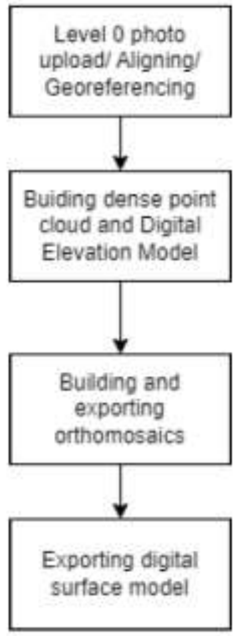

Figure 7 Agisoft Metashape Software Workflow For Generating Canopy Features From Orthomosaics

3.3.2. Training The LSTM Model

Long Short-Term Memory models have historically been used extensively in forecasting crop output [23], [24], commodity price [25], occurrence of pests and diseases [26], and other agronomic applications [27]. Due to its non-linear properties and ability to run parallel input series and discover causal linkages, LSTM models primarily function as predictive models [28].

In this case, we used a multi-layer stacked LSTM to train the model based on the values of the canopy and vegetative indices i.e., CC, CH and EXG recorded during the 2020 cultivation year. The stacked LSTM model had four LSTM layers with 100, 50, 25 and 12 neurons respectively, and one dense layer with a single neuron. A RMSprop optimizer with learning rate of 0.01 was used for training the data. Mean Squared Error (MSE) and MAE were used as the loss function and performance metric of the algorithm respectively. The model was trained 1000 epochs with a batch size of 32 and validation split of 0.2.

3.3.3. Predictions Based On Transfer Learning

Reusing a model that has already been trained on a different problem is known as transfer learning [29]. With transfer learning, a machine can use its understanding of one activity to better generalize about another. There have been many papers on applications of deep transfer learning in agriculture like identification of plant diseases [30], identification system for mildew disease in pearl millet [31], and prediction of soybean crop yields in Brazil [32].

In this paper, a multi-layer stacked LSTM model was trained on the entire 2020 cultivation data. The last few layers were fine-tuned using initial cultivation data of 2021. This model was then used to predict the corresponding canopy and vegetative indices for the post training cultivation time of 2021. This was repeated for the year 2022 too.

The fine-tuned LSTM model was used to predict canopy and vegetative indices for the post training part of 2021 and 2022 cultivation years. Fig. 7, Fig. 8., and Fig. 9, Fig. 10., Fig. 11. were used to depict real and predicted Canopy Cover, Canopy Height, and Excess Green Index values for the remaining cultivation period. The validation accuracies recorded for each of the canopy and vegetative indices have been shown in Table 1.

Table 1 Validation Accuracies Recorded For CC, CH And EXG For The Year 2021
Reprinted with permission from "A Deep Transfer Learning based approach for forecasting spatio-temporal features to maximize yield in cotton crops" by K. C. Gadepally, S. B. Dhal, M. Bhandari, J. Landivar, S. Kalafatis and K. Nowka, 2023 57th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 2023, pp. 1-4, doi: 10.1109/CISS56502.2023.10089748.

Sl. No.	Number of weeks after cultivation in the testing set	Canopy Cover	Canopy Height	Excess Green Index
1	4	93	90	84
2	5	93.2	90.12	84.2
3	6	93.25	90.14	84.28
4	7	93.8	90.38	84.6
5	8	93.92	90.56	84.78
6	9	94.1	90.78	85.35
7	10	94.28	91.2	85.8

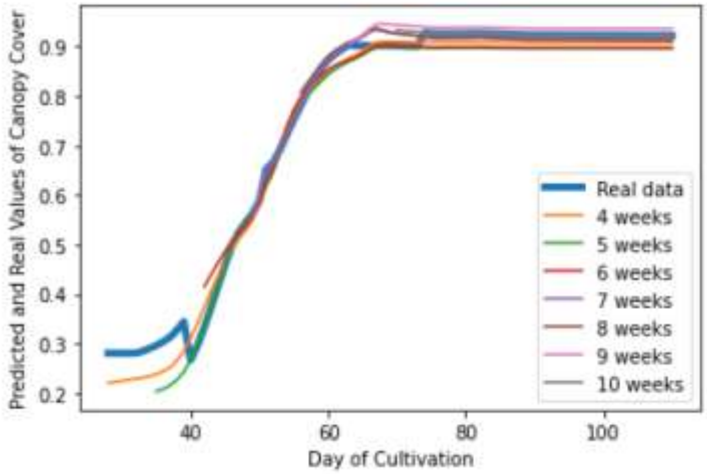

Figure 8 Predicted Vs Real Values Of Canopy Cover For 2021
Reprinted with permission from "A Deep Transfer Learning based approach for forecasting spatio-temporal features to maximize yield in cotton crops" by K. C. Gadepally, S. B. Dhal, M. Bhandari, J. Landivar, S. Kalafatis and K. Nowka, 2023 57th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 2023, pp. 1-4, doi: 10.1109/CISS56502.2023.10089748.

From Figure 8, it was observed that as the training time of LSTM increases, predicted values of Canopy Cover converged to the real values, with accuracy values going from 93 to 94.28.

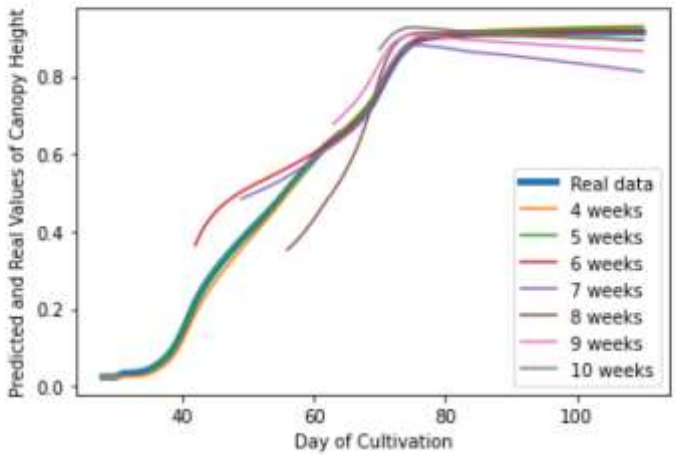

Figure 9 Predicted Vs Real Values Of Canopy Height For 2021
Reprinted with permission from "A Deep Transfer Learning based approach for forecasting spatio-temporal features to maximize yield in cotton crops" by K. C. Gadepally, S. B. Dhal, M. Bhandari, J. Landivar, S. Kalafatis and K. Nowka, 2023 57th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 2023, pp. 1-4, doi: 10.1109/CISS56502.2023.10089748.

From Figure 9, it can be observed that the accuracy values of Canopy Height started at 90 in the case of training time being 4 weeks and improved to 91.2 as the training time reached 10 weeks.

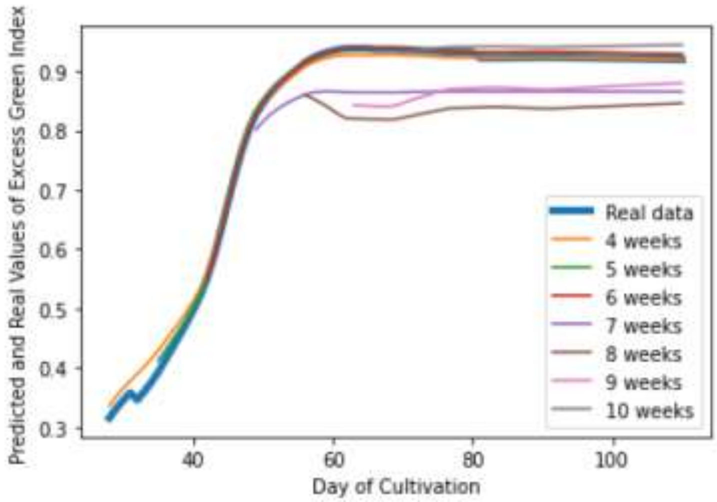

Figure 10 Predicted Vs Real Values Of Excess Green Index For 2021
Reprinted with permission from "A Deep Transfer Learning based approach for forecasting spatio-temporal features to maximize yield in cotton crops" by K. C. Gadepally, S. B. Dhal, M. Bhandari, J. Landivar, S. Kalafatis and K. Nowka, 2023 57th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 2023, pp. 1-4, doi: 10.1109/CISS56502.2023.10089748.

From Figure 10, it was observed that as the training time of LSTM increases, predicted values of Excess Green Index converged to the real values, with accuracy values going from 84 to 85.8.

The fine-tuned LSTM model was also used to predict canopy and vegetative indices for the post training part of 2022 cultivation year. Fig. 10, and Fig. 11 were used to depict real and predicted Canopy Cover and Excess Green Index values for the remaining cultivation period. The validation accuracies recorded for each of the canopy and vegetative indices have been shown in Table 2.

Table 2 Validation Accuracies Recorded For CC And EXG For The Year 2021 Reprinted with permission from "A Deep Transfer Learning based approach for forecasting spatio-temporal features to maximize yield in cotton crops" by K. C. Gadepally, S. B. Dhal, M. Bhandari, J. Landivar, S. Kalafatis and K. Nowka, 2023 57th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 2023, pp. 1-4, doi: 10.1109/CISS56502.2023.10089748.

Sl. No.	Number of weeks after cultivation in the testing set	Canopy Cover	Excess Green Index
1	4	94.1	85.2
2	5	94.25	85.45
3	6	94.8	85.67
4	7	95.3	85.8
5	8	95.43	86.1
6	9	95.89	86.2
7	10	95.97	86.5

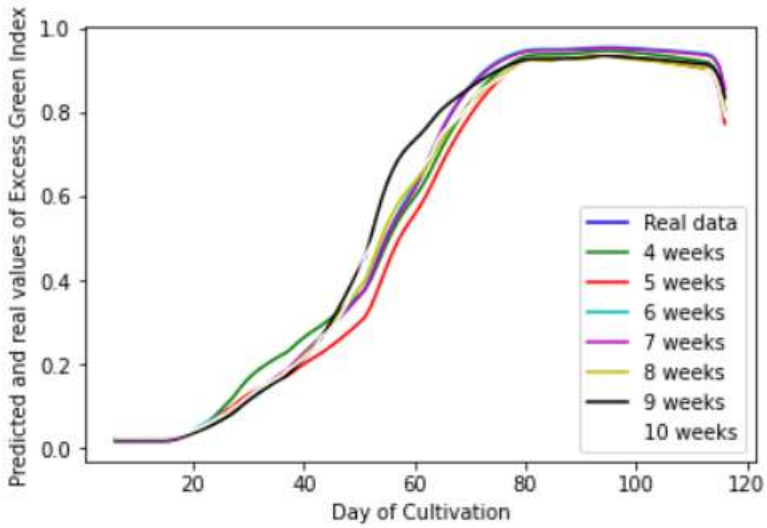

Figure 11 Predicted Vs Real Values Of Canopy Cover For 2022
Reprinted with permission from "A Deep Transfer Learning based approach for forecasting spatio-temporal features to maximize yield in cotton crops" by K. C. Gadepally, S. B. Dhal, M. Bhandari, J. Landivar, S. Kalafatis and K. Nowka, 2023 57th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 2023, pp. 1-4, doi: 10.1109/CISS56502.2023.10089748.

From Figure 11, it was observed that as the training time of LSTM increases, predicted values of Canopy Cover converged to the real values, with accuracy values going from 94.1 to 95.28.

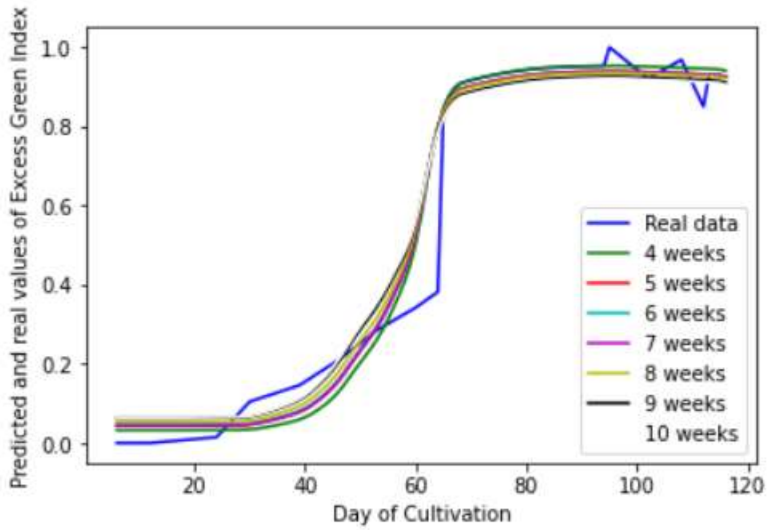

Figure 12 Predicted Vs Real Values Of Excess Green Index For 2022
Reprinted with permission from "A Deep Transfer Learning based approach for forecasting spatio-temporal features to maximize yield in cotton crops" by K. C. Gadepally, S. B. Dhal, M. Bhandari, J. Landivar, S. Kalafatis and K. Nowka, 2023 57th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 2023, pp. 1-4, doi: 10.1109/CISS56502.2023.10089748.

From Figure 12, it was observed that as the training time of LSTM increases, predicted values of Excess Green Index converged to the real values, with accuracy values going from 85.2 to 86.5.

4. SUMMARY[*]

Approaches and methods for better efficiency and accuracy in computer vision applications are the subject of this study. Studies on hardness predictor and image complexity dealt with improving efficiency of computer vision algorithms. On the other hand, improving agricultural drone image-based predictive spatiotemporal models for early season yield prediction through the use of transfer learning on a multilayer, stacked LSTM neural network model was related to enhancing accuracy.

For regression and semantic segmentation, the hardness predictor is effective at assigning hardness values to test images. As the cut off for hardness values increases, harder images are taken as inputs by the computer vision module. As a result, a general decrease in the performance is observed. Further study could be done on designing a common convex loss function for both computer vision application and hardness predictor.

In section 3.3, CNN 2 was trained based on this training data to predict the convolutional layer at which the input image produced one of the strongest activations had a test accuracy of 62 percent. It had a test accuracy of 62 percent. Those results were used to train, validate, and test the convolutional neural network mentioned in the

---

[*] Parts of this chapter are reprinted with permission from "A Deep Transfer Learning based approach for forecasting spatio-temporal features to maximize yield in cotton crops" by K. C. Gadepally, S. B. Dhal, M. Bhandari, J. Landivar, S. Kalafatis and K. Nowka, 2023 57th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 2023, pp. 1-4, doi: 10.1109/CISS56502.2023.10089748.

previous section (CNN 3). It had a test accuracy equal to 99.2 percent, marginally higher than 98 percent when a convolutional neural network was used to train, validate, and test the MNIST dataset without any caveats (CNN 1). This result was validated using the CIFAR-10 dataset too. There was a slight fall in performance by 1.1 percent when CNN 3 was used instead of CNN 1. This could be supplemented by research on CNNs with max pooling layers present.

In the Deep Transfer Learning study, we trained a multi-layer stacked LSTM model using data on canopy and vegetative features, including canopy cover, canopy height, and excess green index, which were obtained in the year 2020. In order to forecast the canopy and vegetative features from the fourth week until the tenth week of harvesting season, the weights of the final few layers of the trained LSTM model were fine-tuned based on the 2021 and 2022 cultivation years' canopy and vegetative index data. This was done using a Deep Transfer Learning technique using a multi-layer stacked LSTM model. As expected, accuracy increased with increase in the time used for fine-tuning of LSTM for all three canopy and vegetative indices. In the case of Canopy Cover, Canopy Height and Excess Green Index, validation accuracy varied between 93 - 95.28 %, 90 - 91.2 % and 84 – 86.5 % respectively.

REFERENCES

[1] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional neural networks. In ECCV, 2014.

[2] Bozinovski, Stevo. "Reminder of the first paper on transfer learning in neural networks, 1976." Informatica 44.3 (2020).

[3] Zhang, Peng, et al. "Predicting failures of vision systems." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.

[4] Daftry, Shreyansh, et al. "Introspective perception: Learning to predict failures in vision systems." 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2016.

[5] Wang, Pei, and Nuno Vasconcelos. "Towards realistic predictors." Proceedings of the European Conference on Computer Vision (ECCV). 2018.

[6] M. Zeiler, D. Krishnan, G. Taylor, and R. Fergus, “Deconvolutional Networks,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2010.

[7] Meyer, L., & Dew, T. (2022). Cotton and wool outlook: May 2022, USDA, Economic Research Service Publication CWS-22e, 1-8. Retrieved from: https://downloads.usda.library.cornell.edu/usdaesmis/files/n870zq801/9593w119d/4742bd79p/CWS-22e.pdf

[8] Zhao, D., Reddy, K. R., Kakani, V. G., Read, J. J., & Koti, S. (2007). Canopy reflectance in cotton for growth assessment and lint yield prediction. European Journal of Agronomy, 26(3), 335-344.

[9] Verdouw, C., Tekinerdogan, B., Beulens, A., Wolfert, S., 2021. Digital twins in smart farming. Agric. Syst. 189, 103046.

[10] Pylianidis, C., Osinga, S., Athanasiadis, I.N., 2021. Introducing digital twins to agriculture. Comput. Electron. Agric. 184, 105942.

[11] Kaul, Monisha, Robert L. Hill, and Charles Walthall. "Artificial neural networks for corn and soybean yield prediction." Agricultural Systems 85.1 (2005): 1-18.

[12] Fernandes, Jeferson Lobato, Nelson Francisco Favilla Ebecken, and Júlio César Dalla Mora Esquerdo. "Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble." International Journal of Remote Sensing 38.16 (2017): 4631-4644.

[13] Ballester, Pedro, and Ricardo Matsumura Araujo. "On the performance of GoogLeNet and AlexNet applied to sketches." Thirtieth AAAI conference on artificial intelligence. 2016.

[14] LeCun, Yann. "LeNet-5, convolutional neural networks." URL: http://yann. lecun. com/exdb/lenet 20.5 (2015): 14.

[15] Zhou, Bolei, et al. "Places: A 10 million image database for scene recognition." IEEE transactions on pattern analysis and machine intelligence 40.6 (2017): 1452-1464.

[16] Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation." IEEE transactions on pattern analysis and machine intelligence 39.12 (2017): 2481-2495.

[17] Parkhi, Omkar M., et al. "Cats and dogs." 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012.

[18] Everingham, Mark, and John Winn. "The pascal visual object classes challenge 2012 (voc2012) development kit." Pattern Analysis, Statistical Modelling

[19] Yann LeCun. The mnist database of handwritten digits. In Technical report, 1998.

[20] Chang, Anjin, et al. "Crop height monitoring with digital imagery from Unmanned Aerial System (UAS)." Computers and Electronics in Agriculture 141 (2017): 232-237.

[21] Amélie Y. Davis et al. "Combined vegetation volume and "greenness" affect urban air temperature". English. In: Applied geography (Sevenoaks) 71 (2016), pp. 106–114.

[22] Suebsombut, P.; Sekhari, A.; Sureephong, P.; Belhi, A.; Bouras, A. Field Data Forecasting Using LSTM and Bi-LSTM Approaches. Appl. Sci. 2021, 11, 11820. https://doi.org/10.3390/app112411820

[23] Jiang, Z., Liu, C., Hendricks, N. P., Ganapathysubramanian, B., Hayes, D. J., & Sarkar, S. (2018). Predicting county level corn yields using deep long short term memory models. arXiv preprint arXiv:1805.12044.

[24] Dhal, Sambandh Bhusan, et al. "Machine Learning Based Models to Forecast in-Season Cotton Growth." ASA, CSSA, SSSA International Annual Meeting. ASA-CSSA-SSSA, 2022.

[25] Yuan, C. Z., & Ling, S. K. (2020, August). Long short-term memory model based agriculture commodity price prediction application. In Proceedings of the 2020 2nd International Conference on Information Technology and Computer Communications (pp. 43-49).

[26] Chen, P., Xiao, Q., Zhang, J., Xie, C., & Wang, B. (2020). Occurrence prediction of cotton pests and diseases by bidirectional long short-term memory networks with climate and atmosphere circulation. Computers and Electronics in Agriculture, 176, 105612.

[27] Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture, 177, 105709.

[28] Murugesan, R., Mishra, E., & Krishnan, A. H. (2021). Deep Learning Based Models: Basic LSTM, Bi LSTM, Stacked LSTM, CNN LSTM and Conv LSTM to Forecast Agricultural Commodities Prices.

[29] Bozinovski, Stevo. "Reminder of the first paper on transfer learning in neural networks, 1976." Informatica 44.3 (2020).

[30] Chen, Junde, et al. "Using deep transfer learning for image-based plant disease identification." Computers and Electronics in Agriculture 173 (2020): 105393.

[31] Coulibaly, Solemane, et al. "Deep neural networks with transfer learning in millet crop images." Computers in Industry 108 (2019): 115-120.

[32] Wang, Anna X., et al. "Deep transfer learning for crop yield prediction with remote sensing data." Proceedings of the 1st ACM SIGCAS