Haoyu Niu and YangQuan Chen

Smart Big Data in Digital Agriculture Applications

Acquisition, Advanced Analytics, and Plant Physiology-informed Artificial Intelligence

November 6, 2023

Springer

Preface

In the dynamic realm of digital agriculture, the integration of big data acquisition platforms has sparked both curiosity and enthusiasm among researchers and agricultural practitioners. This book embarks on a journey to explore the intersection of artificial intelligence and agriculture, focusing on small unmanned aerial vehicles (UAVs), unmanned ground vehicles (UGVs), edge-AI sensors and the profound impact they have on digital agriculture, particularly in the context of heterogeneous crops, such as walnut, pomegranate, cotton, etc. For instance, lightweight sensors mounted on UAVs, including multispectral and thermal infrared cameras, serve as invaluable tools for capturing high-resolution images. Their enhanced temporal and spatial resolutions, coupled with cost-effectiveness and near real-time data acquisition, position UAVs as an optimal platform for mapping and monitoring crop variability across vast expanses. This combination of data acquisition platforms and advanced analytics generates substantial datasets, necessitating a deep understanding of fractional-order thinking, which is imperative due to the inherent "complexity" and consequent variability within the agricultural process. Much optimism is vested in the field of artificial intelligence, such as machine learning (ML) and computer vision (CV), where the efficient utilization of big data to make it "smart" is of paramount importance in agricultural research. Central to this learning process lies the intricate relationship between plant physiology and optimization methods. The key to the learning process is the plant physiology and optimization method. Crafting an efficient optimization method raises three pivotal questions: 1.) What represents the best approach to optimization? 2.) How can we achieve a more optimal optimization? 3.) Is it possible to demand "more optimal machine learning," exemplified by deep learning, while minimizing the need for extensive labeled data for digital agriculture?

In this book, the authors have explored the foundations of the plant physiology-informed machine learning (PPIML) and the principle of tail matching (POTM) framework. They elucidated their role in modeling, analyzing, designing, and managing complex systems based on the big data in digital agriculture. Plant physiology embodies the intricacies of growth, and within this complex system, deterministic and stochastic dynamic processes coexist, influenced by external driving processes characterized and modeled using fractional calculus-based models. These insights better inform the development of complexity-informed machine learning (CIML) algorithms. To practically illustrate the application of these principles, data acquisition platforms, including low-cost UAVs, UGVs, and edge-AI sensors, were designed and built to demonstrate their reliability and robustness for remote and proximate sensing in agricultural applications. Research findings have shown that the PPIML, POTM, CIML, and the data acquisition platforms were reliable, robust, and smart tools for digital agricultural research across diverse scenarios, such as water stress detection, early detection of nematodes, yield estimation, and evapotranspiration (ET) estimation. The utilization of these tools holds the potential to significantly assist researchers and stakeholders in making informed decisions regarding crop management.

Merced, CA, USA, *Haoyu Niu*
November 2023 *YangQuan Chen*

Acknowledgements

The authors would like to thank Dr. Dong Wang for coordinating the USDA project and providing domain knowledge and comments. The authors also would like to thank Dr. Andreas Westphal for coordinating the UC Kearny project and providing the domain knowledge and comments on nematode detection. Thanks go to Prof. Mukesh Singhal, Prof. Wan Du, Dr. Tiebiao Zhao. Their encouraging, critical, and constructive comments and suggestions increased the value of this monograph. Finally, many graduate and undergraduate researchers helped collect and process data: Dong Sen Yan, Stella Zambrzuski, Andreas Anderson, Allan Murillo, Christopher Currier and Joshua Ahmed. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

Contents

Part I Why Big Data Is Not Smart Yet?

1 Introduction . . . 3
1.1 Motivation . . . 3
1.1.1 What Is Smart Big Data in Digital Agriculture? . . . 4
1.1.2 Plant Physiology-informed Artificial Intelligence: A New Frontier . . . 5
1.1.3 Big Data Acquisition and Advanced Analytics . . . 8
1.2 The Book Objectives and Methods . . . 8
1.3 Book Contributions . . . 9
1.4 The Book Outline . . . 10
References . . . 11

2 Why Do Big Data and Machine Learning Entail the Fractional Dynamics? . . . 15
2.1 Fractional Calculus (FC) and Fractional-order Thinking (FOT) . . . 15
2.2 Complexity and Inverse Power Laws (IPLs) . . . 16
2.3 Heavy-tailed Distributions . . . 18
2.3.1 Lévy Distribution . . . 19
2.3.2 Mittag–Leffler PDF . . . 19
2.3.3 Weibull Distribution . . . 20
2.3.4 Cauchy Distribution . . . 20
2.3.5 Pareto Distribution . . . 20
2.3.6 The α-Stable Distribution . . . 21
2.3.7 Mixture Distributions . . . 21
2.4 Big Data, Variability and FC . . . 22
2.4.1 Hurst Parameter, fGn, and fBm . . . 23
2.4.2 Fractional Lower-order Moments (FLOMs) . . . 24
2.4.3 Fractional Autoregressive Integrated Moving Average (FARIMA) and Gegenbauer Autoregressive Moving Average (GARMA) . . . 25

2.4.4 Continuous Time Random Walk (CTRW) 26
2.4.5 Unmanned Aerial Vehicles (UAVs) and Precision Agriculture 26
2.5 Optimal Machine Learning and Optimal Randomness 27
2.5.1 Derivative-free Methods 29
2.5.2 The Gradient-based Methods 29
2.6 What Can the Control Community Offer to ML? 32
2.7 Case Study: Optimal Randomness for Stochastic Configuration Network (SCN) with Heavy-tailed Distributions 38
2.7.1 Introduction 38
2.7.2 SCN with Heavy-tailed PDFs 39
2.7.3 A Regression Model and Parameter Tuning 40
2.7.4 MNIST Handwritten Digit Classification 42
2.8 Chapter Summary 43
References 45

Part II Smart Big Data Acquisition Platforms

3 Small Unmanned Aerial Vehicles (UAVs) and Remote Sensing Payloads 55
3.1 The UAV Platform 55
3.2 Lightweight Sensors 56
3.2.1 RGB Camera 57
3.2.2 Multispectral Camera 57
3.2.3 The Short Wave Infrared Camera 58
3.2.4 Thermal Camera 58
3.3 UAV Image Acquisition and Processing 59
3.3.1 Flight Mission Design 59
3.3.2 UAV Image Processing 60
3.4 Challenges and Opportunities 62
3.4.1 UAVs 62
3.4.2 UAV Path Planning and Image Processing 63
3.4.3 Pre-flight Path Planning 63
3.4.4 Multispectral Image Calibration 63
3.4.5 Thermal Camera Calibration and Image Processing 65
3.4.6 Images Stitching and Orthomosaic Image Generation 66
3.5 Case Study: High Spatial-resolution Has Little Impact on NDVI Mean Value of UAV-based Individual Tree-level Mapping: Evidence from 9 Field Tests and Implications 67
3.5.1 Introduction 67
3.5.2 Material and Methods 68
3.5.3 Results and Discussion 70
3.5.4 Conclusions and Future Work 73
3.6 Chapter Summary 76
References 77

4 The Edge-AI Sensors and Internet of Living Things (IoLT) 81
4.1 Introduction 81
4.2 Proximate Sensors 82
4.2.1 The Spectrometer 82
4.2.2 A Pocket-sized Spectrometer 82
4.2.3 A Microwave Radio Frequency 3D Sensor 84
4.3 Case Study: Onion Irrigation Treatment Inference Using A Low-cost Edge-AI Sensor 85
4.3.1 Introduction 85
4.3.2 Material and Methods 87
4.3.3 Results and Discussion 90
4.3.4 Conclusions and Future Work 93
4.4 Chapter Summary 93
References 94

5 The Unmanned Ground Vehicles (UGVs) for Precision Agriculture .. 97
5.1 Introduction 97
5.2 UGV as Data Acquisition Platform 98
5.2.1 Fundamental Research Questions 99
5.2.2 Low Barriers to Entry 99
5.2.3 Cognitive Algorithms by Deep Learning 101
5.2.4 Swarming Mechanism of UGVs 102
5.3 Case Study: Build A UGV platform for Agricultural Research from A Low-cost Toy Vehicle 103
5.3.1 Introduction 103
5.4 Chapter Summary 104
References 105

Part III Advanced Big Data Analytics, Plant Physiology-informed Machine Learning, and Fractional-order Thinking

6 Fundamentals of Big Data, Machine Learning, and Computer Vision Workflow 111
6.1 Introduction 111
6.2 A Fundamental Tutorial: The Cotton Water Stress Classification with CNN 113
6.2.1 Data Loading 113
6.2.2 Data Pre-processing 114
6.2.3 Train & Test Split 118
6.2.4 Creating the Model 120
6.2.5 The Model Performance Evaluation 125
6.3 Chapter Summary 126
References 126

7 A Low-cost Proximate Sensing Method for Early Detection of Nematodes in Walnut Using Machine Learning Algorithms 127
7.1 Introduction 127
7.2 Material and Methods 129
7.2.1 Study Area 129
7.2.2 Reflectance Measurements with A Radio Frequency Sensor . 129
7.2.3 Groundtruth Data Collection and Processing 131
7.2.4 Scikit-learn Classification Algorithms 131
7.2.5 Deep Neural Networks (DNNs) and TensorFlow 133
7.3 Results and Discussion 133
7.3.1 Data Visualization (Project 45, 2019) 133
7.3.2 Performance of Classifiers (Project 45, 2019) 135
7.3.3 Performance of Classifiers (Project 45, 2020) 138
7.4 Chapter Summary 143
References 143

8 Tree-level Evapotranspiration Estimation of Pomegranate Trees Using Lysimeter and UAV Multispectral Imagery 147
8.1 Introduction 147
8.2 Material and Methods 151
8.2.1 Study Site Description 151
8.2.2 UAV Image Collection and Processing 152
8.3 Results and Discussion 153
8.3.1 Determination of Individual-tree K_c from NDVI 153
8.3.2 The Spatial Variability Mapping of K_c and ET_c 155
8.3.3 Performance of the Individual Tree-level ET Estimation 155
8.4 Conclusions 155
8.5 Chapter Summary 157
References 158

9 Individual Tree-level Water Status Inference Using High-resolution UAV Thermal Imagery and Complexity-informed Machine Learning 163
9.1 Introduction 163
9.2 Material and Methods 164
9.2.1 Experimental Site and Irrigation Management 164
9.2.2 Ground Truth: Infrared Canopy and Air Temperature 165
9.2.3 Thermal Infrared Remote Sensing Data 166
9.2.4 Complexity-informed Machine Learning (CIML) 167
9.2.5 Principle of Tail Matching 168
9.2.6 Machine Learning Classification Algorithms 169
9.3 Results and Discussion 171
9.3.1 Comparison of Canopy Temperature Per Tree Based on Ground Truth and UAV Thermal Imagery 171
9.3.2 The Relationship Between ΔT and Irrigation Treatment 171

9.3.3 The Classification Performance of CIML on Irrigation Treatment Levels . . . 171
9.4 Chapter Summary . . . 174
References . . . 175

10 Scale-aware Pomegranate Yield Prediction Using UAV Imagery and Machine Learning . . . 179
10.1 Introduction . . . 179
10.2 Material and Methods . . . 181
10.2.1 Experimental Field and Ground Data Collection . . . 181
10.2.2 UAV Platform and Imagery Data Acquisition . . . 181
10.2.3 UAV Image Feature Extraction . . . 183
10.2.4 The Machine Learning Methods . . . 185
10.3 Results and Discussion . . . 186
10.3.1 The Pomegranate Yield Performance in 2019 . . . 186
10.3.2 The Correlation between the Image Features and Pomegranate Yield . . . 187
10.3.3 The ML Algorithm Performance on Yield Estimation . . . 187
10.4 Chapter Summary . . . 189
References . . . 192

Part IV Towards Smart Big Data in Digital Agriculture

11 Intelligent Bugs Mapping and Wiping (iBMW): An Affordable Robot-driven Robot for Farmers . . . 197
11.1 Introduction . . . 197
11.2 Existing Solutions . . . 198
11.3 iBMW Innovation . . . 199
11.3.1 Cognitive of Pest Population Mapping and Wiping . . . 200
11.3.2 iBMW with TurtleBot 3 as "Brain" . . . 202
11.3.3 Real-time Vision Processing . . . 202
11.3.4 Optimal Path Planning Enabled by iBMW . . . 204
11.3.5 Ethical, Cultural and Legal Matters . . . 204
11.4 Measuring Success . . . 205
11.4.1 NOW Population Temporal and Spatial Distribution . . . 205
11.4.2 The Amount of Pesticide Being Used . . . 205
11.4.3 The Target Trees Almond Yield . . . 206
11.5 Chapter Summary . . . 206
References . . . 206

12 A Non-invasive Stem Water Potential Monitoring Method Using Proximate Sensor and Machine Learning Classification Algorithms . . . 209
12.1 Introduction . . . 209
12.2 Material and Methods . . . 210
12.2.1 Walnut Study Area . . . 210
12.2.2 Reflectance Measurements with A Radio Frequency Sensor . 211

12.2.3 Data Collection and Processing 211
12.2.4 Scikit-learn Classification Algorithms 211
12.3 Results and Discussion 212
12.4 Chapter Summary 213
References 215

13 A Low-cost Soil Moisture Monitoring Method by Using Walabot and Machine Learning Algorithms 217
13.1 Introduction 217
13.2 Material and Methods 218
13.2.1 Study Site 218
13.2.2 The Proximate Sensor 219
13.2.3 Experiment Setup 219
13.2.4 Data Collection and Processing 219
13.3 Results and Discussion 222
13.3.1 Linear Discriminant Analysis Performance 223
13.3.2 Principle Component Analysis Performance 224
13.4 Chapter Summary 225
References 226

14 Conclusions and Future Research 229
14.1 Concluding Remarks 229
14.2 Future Research Towards Smart Big Data in Digital Agricultural Applications 230

Index 233

List of Figures

1.1 The ML can be classified as supervised, unsupervised, semi-supervised, and Reinforcement Learning (RL) based on whether or not human supervision is included. According to whether or not the ML algorithms can learn incrementally on the fly, they can be classified into online and batch learning. Based on whether or not the ML algorithms detect the training data patterns and create a predictive model, the ML can be classified into instance-based and model-based learning. 6

2.1 Inverse power law (complexity "bow tie"): On the left are the systems of interest that are thought to be complex. In the center panel, an aspect of the empirical data is characterized by an inverse power law (IPL). The right panel lists the potential properties associated with systems with data that have been processed and yield an IPL property. See text for more details. 17
2.2 Complex signals (IPL): Here, the signal generated by a complex system is depicted. Exemplars of the systems are given as the potential properties arising from the systems' complexity. 18
2.3 Normalized difference vegetation index (NDVI) mapping of pomegranate trees. 27
2.4 Data analysis in nature. 28
2.5 The 2-D Alpine function for derivative-free methods. 29
2.6 Sample paths. Wei et al. investigated the optimal randomness in a swarm-based search. Four heavy-tailed PDFs were used for sample path analysis. The long steps, referring to the jump length, frequently happened for all distributions, which showed strong heavy-tailed performance. 30
2.7 Gradient descent and its variants. 31
2.8 The integrator model (embedded in $G(z)$). The integrator in the forward loop eliminates the tracking steady-state error for a constant reference signal (internal model principle (IMP)). 32

2.9 Training loss (**left**); test accuracy (**right**). It is obvious that for different zeros and poles, the performance of the algorithms is different. One finds that both the $b = -0.25$ and $b = -0.5$ cases perform better than does the stochastic gradient descent (SGD) momentum. Additionally, both $b = 0.25$ and $b = 0.5$ perform worse. It is also shown that an additional zero can improve the performance, if adjusted carefully. (Courtesy of Professor Yuquan Chen) 35
2.10 Training loss (**left**); test accuracy (**right**). (Courtesy of Professor Yuquan Chen) 37
2.11 Performance of SCN, SCN-Lévy, SCN-Weibull, SCN-Cauchy and SCN-Mixture. The parameter L is the hidden node number. 42
2.12 The handwritten digit dataset example. 43
2.13 Classification performance of SCNs. 44
2.14 Timeline of FC (courtesy of Professor Igor Podlubny). 45

3.1 (**a**) The QuestUAV 200 UAV. (**b**) The MK Okto XL 6S12. (**c**) The DJI S1000. (**d**) The eBee Classic. (**e**) The Hover. 55
3.2 The Survey 2 sensors and the reflectance calibration ground target package. 57
3.3 The Rededge M is a commonly used multispectral camera. The Rededge M has five bands, which are Blue, Green, Red, Near-infrared, and Red edge. It has a spectral resolution of 1280 $\times$ 960 pixel, with a 46° field of view. 58
3.4 The SWIR 640 P-Series, which is a shortwave infrared camera, has also been commonly used for agricultural research. The spectral band is from 0.9 μm to 1.7 μm. The accuracy for the SWIR camera is $\pm$ 1 °C. It has a resolution of 640 $\times$ 512 pixels. 59
3.5 The thermal camera has a resolution of 640 $\times$ 480 pixels. The spectral band is from 7 μm to 14 μm. The dimension of the thermal camera is 34 mm $\times$ 30 mm $\times$ 34 mm. The accuracy is designed to be $\pm$ 2 °C. 60
3.6 The user interface of Mission Planner. The example of flight mission was for nematode data collection using UAV for Project 30 at UC Kearny Center. See Chapter 7 for more details. 61
3.7 Agisoft Metashape image processing workflow: (**a**) Align Photos. (**b**) Build Mesh. (**c**) Generate orthomosaick. 62
3.8 The UAV image reflectance calibration. (**a**) A color panel. (**b**) The MAPIR target calibration board. (**c**) The Rededge M calibration board. 64
3.9 The pomegranate study site at USDA. The pomegranate was planted in 2010 with a 5 m spacing between rows and 2.75 m within-row tree spacing in a 1.3 ha field. There are two large weighing lysimeters located in the center of the field, marked in red boxes. 68

3.10 The mean NDVI values of each sampling tree at 60m, 90m, and 120m on May 8^{th}, 2019. The x-axis was the identification number (ID) for sampling trees, 82 in total. The y-axis was the mean NDVI value for each tree canopy. 70
3.11 The mean NDVI values of each sampling tree at 60m, 90m, and 120m on Sep 19^{th}, 2019. The x-axis was the identification number (ID) for sampling trees, 82 in total. The y-axis was the mean NDVI value for each tree canopy. 71
3.12 The mean NDVI values of each sampling tree at 60m, 90m, and 120m on Oct 3^{rd}, 2019. The x-axis was the identification number (ID) for sampling trees, 82 in total. The y-axis was the mean NDVI value for each tree canopy. 71
3.13 The individual tree-level mean NDVI correlation between 120 m and 90 m on Oct 3^{rd}, 2019. The x-axis was the mean NDVI values for sampling trees at 120 m flight height. The y-axis was the mean NDVI values for sampling trees at 90 m flight height. 72
3.14 The NDVI distribution of two individual lysimeter trees at 60m, 90m and 120m on May 8^{th}, 2019. 73
3.15 The NDVI distribution of two individual lysimeter trees at 60m, 90m and 120m on Sep 19^{th}, 2019. 74
3.16 The NDVI distribution of two individual lysimeter trees at 60m, 90m and 120m on Oct 3^{rd}, 2019. 75

4.1 The spectrometer EPP2000-VIS-50. The StellarNet EPP2000 Spectrometers are miniature fiber optic instruments for ultraviolet (UV), visisble (VIS), and near-infrared (NIR) measurements in 190 - 1700 nm ranges. 82
4.2 A light and small near-infrared spectrometer, Scio. The low power consumption and zero warm up time make it highly responsive and extremely efficient, which allows it to perform hundreds of samplings from a small rechargeable battery. The Scio spectrometer works in NIR at wavelengths of 700 - 1100 nm. It can also be integrated into smart phone as an Edge-AI device. 83
4.3 A radio frequency 3D sensor, Walabot Developer (Front and back view). The frequency range was 3.3 - 10 GHz for the US/FCC model and 6.3 - 8 GHz for the EU/CE mode. The average transmitted energy of both models was below 41dBm/MHz and did not constitute health concerns for the user, nor did it impact the tested tissues. 85
4.4 The princile of Walabot. The sensor uses an antenna array to send radio frequency to the area in front of it, e.g. walnut leaves, and then captures the returning signals (reflectance or response). The citation signals are produced, and the reflectance is then recorded by the radio frequency integrated circuits (RFIC). 85

4.5 The measurement signal of the radio frequencies reflectance of the walnut leaves using Walabot. The research hypothesis is that the reflectance of walnut leaves will be significantly different towards root-lesion nematode infection levels. The spectral curve obtained from the leaves of walnut trees infected with nematodes could show the characteristic pattern of walnut under nematode stress. For more details of this research work, please refer to Chapter 7. 86
4.6 The onion field (image taken by the "Hover" UAV platform). The study field was in the USDA-ARS, San Joaquin Valley Agricultural Sciences Center (36.59 °N, 119.51 °W), Parlier, California. Since 2016, an onion test filed had been set up for research of biomass soil amendments and deficit irrigation. There were three irrigation treatment levels, High, Medium and Low. 87
4.7 The NIR reflectance measurement of SCIO using smart phone. 88
4.8 The reflectance measurement of EPP2000-VIS-50. Each onion sampling was measured three times to reduce the likelihood of errors or anomalous results. 89
4.9 Classifier Comparison (PCA dimension reduction) 91
4.10 Classifier Comparison (LDA dimension reduction) 92

5.1 The toy vehicle. 103
5.2 The UGV platform. 104

6.1 The cotton field experimental layout. 113
6.2 A 5 × 5 grid in the case of displaying 25 cotton images. 116
6.3 Cotton segmentation with morphological image processing methods. **(a)** The original UAV-based RGB image; **(b)** A rough mask image after Otsu's methods; **(c)** The mask after opening and dilation operations; **(d)** The cotton canopy image with the mask in **(c)**. 117
6.4 A visual demonstration of an image dataset after morphological image processing. 119
6.5 The model summary. 123
6.6 The training and testing accuracy of the CNN model. 124
6.7 The summary of prediction results of the CNN model for original image dataset at different sampling days. The irrigation treatments were labeled as follows: "rainfed" as "0", "fully irrigated" as "1", "percent deficit" as "2", and "time delay" as "3". The color bar on the right side represented the number of test samples. 125

7.1 The study was conducted in an experimental walnut orchard at the Kearney Agricultural Research and Extension Center, 9240 S Riverbend Ave, Parlier, CA, 93648, USA (36.595°N, 119.508°W). There were three replications to detect the nematode infection levels, Project 30 (a, b), Project 45, and Project Smith. 130

7.2 The measurement signal of the radio frequencies reflectance of the walnut leaves using Walabot. The research hypothesis is that the reflectance of walnut leaves will be significantly different towards root-lesion nematode infection levels. The spectral curve obtained from the leaves of walnut trees infected with nematodes could show the characteristic pattern of walnut under nematode stress. 131
7.3 The Linear Discriminant Analysis (LDA) was used for data visualization, which learned the most critical axes between the classes. The axes were then used to define hyperplane to project the high-dimensional training data into two dimensions, which gained important insight by visually detecting patterns. The *x*-axis and *y*-axis had no scale because of hyperplane projection. The nematode infection levels were successfully clustered into four levels. The color bar values 1, 2, 3, and 4 represented the four nematode infection levels from 1 to 4 (low to high) (Table 7.2). 134
7.4 The Linear Discriminant Analysis (LDA) was used for three-dimensional data visualization. The *x*-axis and *y*-axis had no scale because of hyperplane projection. The nematode infection levels were successfully clustered into four levels. The clusters' color was related to Fig. 7.3. 135
7.5 A comparison of six classifiers in Scikit-learn on nematode testing dataset (Project 45, 2019). A confusion matrix was a summary of prediction results on a classification problem. The number of correct and incorrect predictions were tallied with count values and divided into classes. The confusion matrix provided an insight not only into the errors being made by a classifier but more importantly the types of errors that were being made. "True label" meant the ground truth of nematode infestation levels. "Predicted label" identified the nematode infestation levels predicted by the trained model. 137
7.6 There were k models (in this case, $k = 100$). The best accuracy could be as high as 0.9 (marked as a red dot). The lowest accuracy was around 0.58 when the training instances could not represent the nematode infection levels dataset very well. The mean accuracy was 0.82 for the DNN classier model. 138
7.7 A comparison of six classifiers in Scikit-learn on nematode testing dataset (Project 45, 2020). Table 7.8 showed the details of the test process of the *k*-nearest neighbors method. 141
7.8 The test performance of DNN for Project 45, 2020. There were 100 models and the best accuracy could be as high as 0.88 (marked as a red dot). The lowest accuracy was around 0.4 when the training instances could not represent the nematode infection levels dataset very well. The mean accuracy was 0.7 for the DNN classier model. .. 142

8.1 The Pomegranate study site. Field studies were conducted in an experimental pomegranate (*Punica granatum L.*, cv 'Wonderful') field at the USDA-ARS, San Joaquin Valley Agricultural Sciences Center (36.594 °N, 119.512 °W), Parlier, California, 93648, USA. The pomegranate field was randomly designed into 16 equal blocks, with four replications, to test four irrigation levels. The irrigation volumes are 35%, 50%, 75%, and 100% of ET_c, which was measured by the weighing lysimeter in the field. There were five sampling trees in each block, 80 sampling trees in total, marked with red labels. 152
8.2 Seasonal K_c and NDVI for the tree in lysimeter. The sampling data started on May 8^{th} and ended on Oct 29^{th}. The linear relationship had an intercept of 0.6114 and a slope of 1.6493. It also had a high correlation coefficient of 0.8865, indicating a significant correlation between the K_c and NDVI at the individual-tree level during the growing season in 2019. 154
8.3 The K_c and ET_c maps of May 8^{th}, 2019. 156
8.4 100% ET irrigation sampling trees vs lysimeter tree. To validate the linear regression model on individual-tree-level ET estimation, the author compared the 20 sampling trees with the lysimeter daily ET_c for the UAV flight dates (almost the whole growing season). The trends of the daily ET_c for the 20 sampling trees (100% irrigation) and the lysimeter tree were shown. Each 'Serie' meant an individual tree in the field. 157
8.5 The boxplot of 100% ET sampling trees vs lysimeter tree. Compared with the lysimeter tree, the linear regression model estimated tree-level ET with an R^2 and mean absolute error (MAE) of 0.9143 and 0.39 mm/day, respectively. 158

9.1 The pomegranate field at the USDA-ARS (36.594°N, 119.512°W). The weighing lysimeter is located in the center of the pomegranate field, marked as a red box. The blue marks are where the 14 IRT sensors were installed. 165
9.2 The IRT sensor was installed 4.5 m above the soil surface, with a FOV of 20°. A quadcopter was used as the low-cost UAV platform (less than $1000) to collect high-resolution thermal images at the height of 60 m. 166
9.3 **(A)** All of the UAV thermal images were stitched together to generate the orthomosaick images in Metashape (Agisoft LLC, Russian); **(B)** To obtain the individual tree canopy images of the 250 sampling trees, the author used the SVM for classifying the tree canopy. **(C)** Histogram was generated for each tree canopy to check the variability of each tree canopy temperature. 168

9.4 The correlation between the canopy temperature per tree measured by the IRT sensors and UAV thermal camera. The coefficient of determination (R^2) was 0.8668, which indicated that the difference between the ground truth and UAV thermal camera was acceptable. The method was reliable for monitoring tree-level canopy temperature. 170
9.5 The ΔT was significantly higher in the 35% irrigation treatment than the 100% irrigation treatment on different days. The values of ΔT decreased as the irrigation increased. This finding emphasized the importance of irrigation on the tree canopy temperature response. 172
9.6 The summary of prediction results using histogram information on the tree-level irrigation treatment inference. "True label" meant the ground truth of ET_c based irrigation treatment levels. "Predicted label" identified the irrigation treatment levels predicted by the trained model. To simplify the visualization, 30% and 50% ET irrigation were labeled as "0", denoting low-level irrigation; 75% and 100% ET irrigation were labeled as "1", which meant high-level irrigation. 173
9.7 The test performance for the histogram dataset. The t-distributed stochastic neighbor embedding (TSNE) method was used for data visualization, which learned the most critical axes between the classes. The axes were then used to define the hyperplane to project the high-dimensional training data into two dimensions, which gained important insight by visually detecting patterns. The *x*-axis and *y*-axis had no scale because of hyperplane projection. The irrigation treatment levels were successfully clustered into low-level (blue) and high-level (green). 175

10.1 The pomegranate field was randomly designed into 16 equal blocks, with four replications, to test four irrigation levels. The irrigation volumes are 35%, 50%, 75%, and 100% of ET_c, which was measured by the weighing lysimeter in the field. 182
10.2 There were five sampling trees in each block, 80 sampling trees in total, marked with red labels. 182
10.3 The pomegranate yield performance at the individual tree level in 2019. For the 35% irrigation treatment, the total fruit weight per tree was 23.92 kg, which produced the lowest yield. For the 50% irrigation treatment, the total fruit weight per tree was 27.63 kg. For 75% and 100% irrigation treatment, the total fruit weight per tree was 29.84 kg and 34.85 kg, respectively. 186
10.4 The correlation between the vegetation indices and yield. 188

10.5 The "Decision Trees" method training process. The "Decision Trees" usually uses a white box model, which means the explanation for the condition is easily explained by Boolean logic if a given situation is observable in a model. As shown here, the "Decision Trees" ML model started at the root node, if the NDVIre value were less than 0.334, the prediction process would move to the leaf child node. In this case, the model would predict that the input was a low-yield pomegranate tree. A node's gini attribute measures its impurity: a node is "pure (gini = 0)" if all the training instances it applies are from the same class. 190
10.6 The comparison of the eight different ML classifiers on individual tree level yield prediction. "True label" meant the ground truth of the yield. "Predicted label" identified the individual tree yield predicted by the trained model. The value 0 meant the low yield; value 1 meant the high yield. 191

11.1 The iBMW workflow. The iBMW is an intelligent robot which can recognize and classify the NOW by using deep learning neural networks. By using swarming mechanism, several iBMWs can scout in orchard together to realize pest population mapping in a large scale. Based on the mapping analysis, the iBMWs can determine which areas are at the greatest risk and whether wiping treatment is needed. 200
11.2 Concept iBMW. 201
11.3 Road detection by Jetson TX2. 203

12.1 This study was conducted in a walnut orchard at Merced, CA, 95340, USA (37.47°N, 120.45°W). There were three types of spacing treatment, 22' × 22', 26' × 26', and 30' × 30'. The region of interest was marked with the red box. To determine the SWP of a walnut tree, midday SWP was measured by using a pressure chamber weekly from May to September 2018. 210
12.2 A comparison of six classifiers in Scikit-learn on SWP prediction. (a) "Decision Trees"; (b) "Nearest Neighbors"; (c) "Neural Net"; (d) "AdaBoost"; (e) "Random Forest"; (f) "Support Vector Machine". A confusion matrix was a summary of prediction results on a classification problem. The number of correct and incorrect predictions were tallied with count values and divided into classes. The confusion matrix provided an insight not only into the errors being made by a classifier but more importantly the types of errors that were being made. "True label" meant the ground truth of SWP levels. "Predicted label" identified the SWP levels predicted by the trained model. 214

13.1 Radio Frequency reflectance. The Walabot Developer is a programmable 3D sensor that uses radio frequency to see through the soil and creates a reflectance image within one second. 218
13.2 Walabot data collection. The experiment was conducted in the MESA Lab. The soil was sampled in an almond field near the lab and was divided into 3 cups. 220
13.3 LDA results for soil moisture measurement. There were five different soil moisture levels with different colors, “Dry,” “Wet1,” “Wet2,” “Wet3,” and “WetTotal”, which meant the soil sampling was saturated. LDA classifiers firstly reduced the original dimension to 2 components. Different colors meant different soil moisture levels and the axes of the figure were dimensionless. The LDA could classify the five different levels of soil moisture in different areas of the coordinate, so that the LDA could classify the soil moisture with an state-of-art performance. 224
13.4 PCA results for soil moisture measurement. “WetTotal” points were on the left and right sides of the image. The “Wet3” and “Wet2” data points did not drop in the same area. The reason might be that the PCA could not detect the features difference from the data. Similar to LDA, the PCA classifiers firstly reduced the original dimension to 2 components. Then, each classifier was tested against reduced dimensionality data with the component as 2. . 226

List of Tables

2.1 The 10 Vs of big data. . . . 22
2.2 Variability in multiple research topics. . . . 23
2.3 General second-order algorithm design. The parameter ρ is the loop forward gain; see text for more details. . . . 34
2.4 General third-order algorithm design, with parameters defined by Equation (2.39). . . . 36
2.5 The continuous time fractional-order system. . . . 36
2.6 SCNs with key parameters. . . . 40
2.7 Performance comparison of SCN models on regression problem. . . . 41
2.8 Performance comparison between SCN, SCN-Lévy and SCN-Cauchy. 43

3.1 The specifications of “Hover”. The quadcopter is equipped with high efficient power system, including T-Motor MN3508 KV380 motor, 1552 folding propeller and Foxtech Multi-Pal 40A OPTP ESC, to ensure long flight time. . . . 56
3.2 Orthomosaic images generation workflow in Agisoft Metashape. . . . 61
3.3 The UAV flight schedule. The UAV flight height was 60 m, 90 m, and 120 m to acquire different high-resolution multispectral imagery. Data was collected successfully for three different days, 5-8-19, 9-19-19, and 10-3-19. . . . 69
3.4 Entropy was used in the quantitative analysis and evaluation of image information, because it provided better comparison of the image details. Higher value of entropy meant more detailed information in the image. . . . 76

4.1 The specifications of spectrometer EPP2000-VIS-50. . . . 83
4.2 The specifications of spectrometer Scio. . . . 84
4.3 Top 5 performing classifiers using PCA and MLP and their grid search parameters. . . . 92

5.1 Specifications of several UGVs and the proposed Personal UGV. . . . 100

7.1 Nematode numbers per gram of root in rootstock genotypes examined by Walabot measurements (Project 45, 2019). “RLN per gram of root” meant the root-lesion nematode number per gram of root extracted in laboratory procedures. 132
7.2 The classification of the nematode infestation levels (Project 45, 2019). 132
7.3 The performance of classification methods (Project 45, 2019). 136
7.4 The *k*-nearest neighbors performance (Project 45, 2019). 136
7.5 The classification of the nematode infestation levels (Project 45, 2020). “RLN per gram of root” meant the root-lesion nematode number per gram of root that was extracted in laboratory procedures. 139
7.6 Nematode numbers per gram of root in rootstock genotypes examined by Walabot measurements (Project 45, 2020). “RLN per gram of root” meant the root-lesion nematode number per gram of root that was extracted in laboratory procedures. “ID” is a short form for “identity”. 139
7.7 The performance of classification methods (Project 45, 2020). 140
7.8 The *k*-nearest neighbors performance (Project 45, 2020). 140

8.1 ET estimation using UAV platforms. 149
8.2 Multispectral and thermal infrared sensors on UAV platforms. 150
8.3 Flight missions at the USDA in 2019. The UAV flight missions were configured by using the MissionPlanner (Ardupilot, USA). The flight altitude was set up as 60 m. The overlapping of UAV images was set up as 80% forward and 70% by the side. 153

9.1 The classification performance of CIML algorithms on irrigation treatment levels at individual tree level. All the methods showed a state-of-art performance, with an overall accuracy of 87%. The “Naive Bayes” had the highest accuracy of 0.90. 174

10.1 UAV image features used in this study. 183
10.2 The “Decision Tree” performance on yield prediction. “NA” stands for “Not available”. 189
10.3 The performance of ML methods on yield prediction. 189

12.1 The stem water potential of walnut leaves. There were 16 sampling trees, which had unique ID numbers. For example, “20-4” meant the fourth tree on row 20. Each sampling tree was measured five times to reduce the likelihood of errors or anomalous results. 212
12.2 The SWP levels of walnut leaves. Based on the SWP range, the walnut trees were classified into three levels. 212

12.3 The performance of classification methods. The accuracy of the "Support Vector Machine" was 0.62. The "Random Forest" also had a low prediction accuracy of 0.60. The "AdaBoost," "Nearest Neighbors," and "Neural Network" had an accuracy of 0.60, 0.65, and 0.62, respectively. 213
12.4 The "Decision Trees" performance on SWP prediction. In the analysis as "Decision Tree" for the SWP Level 1, the trained model predicted the test data was in the range of Level 3. Therefore, the prediction accuracy (F1-score) of Level 1 was zero. The model successfully classified 22 out of 25 samplings in Level 2. For the Level 2, the model had performance with an F1-score of 0.92. For the Level 3, the model classified 9 out of 10 samplings. The overall prediction accuracy of the trained model was 0.78. 215

13.1 Soil samplings. All the soil samplings were from the same spot in the almond field to make sure they are homogeneous. The soil was dried out to make sure all the 3 cups of soil were at the same lowest moisture level. The weights of three cups of dry soil were 632 grams, 630 grams, and 634 grams. 6 g or 8 g water was added in every cup each time (10 times in total) to increase the soil moisture until the soil moisture was saturated. 220
13.2 Classifiers accuracy. The best classifiers were "Nearest Neighbors," "Gaussian Process," "Decision Tree," "Random Forest," "Neural Net," and "Naive Bayes" with an accuracy of 95%. The "QDA" was with 90% accuracy. The "Linear SVM" and "AdaBoost" were worst with 55% accuracy. 222
13.3 The LDA Methods. Several LDA methods were used for soil moisture classification. 223
13.4 The PCA Methods. 225

Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
ARS	Agricultural Sciences Center
BRDF	Bidirectional Reflectance Distribution Function
CIMIS	California Irrigation Management Information System
CIML	Complexity-informed Machine Learning
CNNs	Convolutional Neural Networks
CRP	Calibrated Reflectance Panel
DEM	Digital Elevation Model
DLS	Downwelling Light Sensor
DN	Digital Number
DNNs	Deep Neural Networks
DOY	Day of Year
DTD	Dual Temperature Difference
ET	Evapotranspiration
FOV	Field of View
GPS	Global Positioning System
GPU	Graphics Processing Unit
HRMET	High Resolution Mapping of ET
ID	Identity
IoLT	Internet of Living Things
IR	Infrared
JPG	Joint Photographic Experts Group
LAI	Leaf Area Index
LDA	Linear Discriminant Analysis
MAE	Mean Absolute Error
METRIC	Mapping Evapotranspiration with Internalized Calibration
ML	Machine Learning
MLP	Multi-layer Perceptron
NDVI	Normalized Difference Vegetation Index
NIST	National Institute of Standards and Technology

NIR	Near Infrared
OSEB	One Source Energy Balance
PA	Precision Agriculture
PCA	Principal Component Analysis
PDF	Probability Distribution Function
POTM	Principle of Tail Matching
PPIML	Plant Physiology-informed Machine Learning
QDA	Quadratic Discriminant Analysis
RGB	Red, Green, and Blue
RMSE	Root Mean Square Error
RSEB	Remote Sensing Energy Balance
SCN	Stochastic Configuration Network
SEBAL	Surface Energy Balance Algorithm for Land
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
SWIR	Short-wave Infrared
TIR	Thermal Infrared
TSEB	Two-source Energy Balance
TSEB-PT	Priestley-Taylor TSEB
UAVs	Unmanned Aerial Vehicles
UGVs	Unmanned Ground Vehicles
US	United States
USDA	United States Department of Agriculture
VIS	Visable

Part I
Why Big Data Is Not Smart Yet?