DaQual: Data Quality Assessment for Tree Trunk Relative Water Content Sensors in a Pomegranate Orchard

Yue Zhang
yzhang58@ucmerced.edu
University of California, Merced
Merced, USA.

Abdias Tellez Benitez
atellezbenitez@ucmerced.edu
University of California, Merced
Merced, USA.

Reza Ehsani
rehsani@ucmerced.edu
University of California, Merced
Merced, California, USA

Shijia Pan
span24@ucmerced.edu
University of California, Merced
Merced, California, USA

ABSTRACT

High-fidelity sensor data quality is the fundamental base of smart agriculture. Since crop information inference and cultivating strategy optimization mainly depend on data-driven methods; the data quality assessment is essential to ensure the reliability of the IoT systems for smart agriculture. The traditional data quality assessment methods focus on sensor data consistency with the costly reference truth, which is often not scalable for agricultural applications.

We present *DaQual*, a data-driven data quality assessment scheme for tree trunk relative water content sensors in a pomegranate orchard. The objective is to leverage the neural network architecture to learn the underlying relationship between sensors deployed on the same farm and utilize the quantified associated relationship between sensors to assess its reliability. *DaQual* first builds a prediction model for each sensor with all other sensors' data as input. Then, the trained network's parameter values (weights) are used to quantify the contribution of the sensor – the higher the contribution, the higher its data quality. We evaluate *DaQual* via a real-world tree trunk relative water content sensor dataset with nine sensors deployed in a pomegranate orchard, and our scheme demonstrates up to 1.8× improvements when used to select a subset of sensors with high data quality.

CCS CONCEPTS

• **Applied computing** → **Agriculture**.

KEYWORDS

Data Quality, Smart Agriculture, Quality Assessment

ACM Reference Format:

Yue Zhang, Abdias Tellez Benitez, Reza Ehsani, and Shijia Pan. 2022. *DaQual*: Data Quality Assessment for Tree Trunk Relative Water Content Sensors in a Pomegranate Orchard. In *The 20th ACM Conference on Embedded Networked Sensor Systems (SenSys '22), November 6–9, 2022, Boston, MA, USA*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3560905.3568415

SenSys '22, November 6–9, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9886-2/22/11…$15.00
https://doi.org/10.1145/3560905.3568415

1 INTRODUCTION

High quality sensor data is essential for smart agriculture using Internet of things (IoT) systems [31, 50], which provides long-term and continuous monitoring for farms and orchards [6, 15]. However, the data quality collected in fields may vary a lot, due to reasons like manufacturing processes and deployment environment [32, 49]. Low-quality sensing data could result in low inference accuracy, and even lead to erroneous decisions and/or yield deduction for smart farms [23]. Therefore, data quality assessment is essential to select high-quality sensing data to provide more accurate information for crop status estimation and cultivating strategy optimization.

Studies show that data quality is a multi-dimensional concept, including validity, completeness, consistency, uncertainty, and information quality [10, 19, 27, 39, 42]. In this work, we focus on the **information quality**, where we aim to select accurate and informative sensors to ensure the accuracy of inference. One way of assessing the data quality is to conduct a comparison with a reference truth, which is acquired by a more sophisticated and reliable device/system [4]. This could be labor-intensive and time-consuming for large scaled deployment at farms. For example, the reference truth for the tree trunk relative water content sensor is to use a pressure chamber to measure stem water potential leaf by leaf [17, 18], which is not scalable for large-scale deployment. As illustrated in Figure 1, the normalized sensor data (dotted line) and the reference truth (solid line) for sensor $n1$ and $n2$ (blue and red) in eight days demonstrates inconsistent sensor data trends. $n1$ demonstrates a more consistent trend with the reference truth but a larger offset, while $n2$ demonstrate a closer value range with higher discrepancy compared to the reference truth. Therefore, it is difficult to use straightforward approaches, such as Pearson correlation coefficients, to **determine the data quality** of the sensor. Furthermore, it would be even more challenging if the **reference truth is not available** all the time.

In this work, we present *DaQual*, a sensor data quality assessment framework for sensors deployed on the farm. *DaQual* establishes a time sequence prediction model from the tree trunk relative water content sensors in a pomegranate orchard and use the parameters of the trained neural network (weights of each input time sequence) to quantify the data quality of each sensor. The intuition is that the neural network can learn the underlying relationships between sensor data on the same farm (similar environment), and the time sequence that has a higher weight contribution to the prediction of other sequences indicates high data quality. The contributions of this work are as follows:

- We present *DaQual*, a data-driven sensor data quality assessment framework, which quantifies sensor data quality

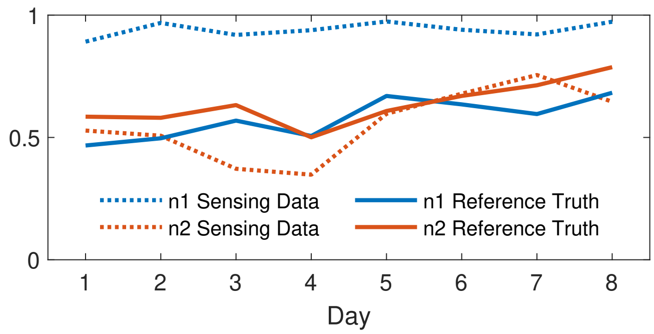

Figure 1: The normalized tree trunk relative water content sensor data and reference truth of stem water potential.

and selects high-quality sensor groups for reliable tree trunk relative water content monitoring.

- We introduce a new data quality assessment scheme via time series prediction and neural network parameter analysis.
- We evaluate this framework via a real-world dataset, where nine sensors measure tree trunk relative water content in a pomegranate orchard.

2 RELATED WORK

We review past research on IoT for agriculture, sensor calibration, and IoT data quality.

2.1 IoT for Agriculture

Agriculture is important for human food security, and many IoT sensing systems are proposed to solve the challenges in agriculture, including plant status monitoring, strategy design, and food supply chain optimization. Plant status monitoring provides a real-time and accurate measurement of plant growth and health, such as plant water status [5, 8, 13], soil humidity [29], and other weather conditions [40, 45]. Accordingly, some studies focus on optimizing strategies for plant growth based on the IoT sensing data, such as irrigation management [22] to save water and preventive strategies in different weather conditions to improve crop yield [45]. IoT technologies are utilized in agriculture to optimize the food supply chain, which helps reduce transport costs and improve food quality. Recently, a more pervasive data collection platform for agriculture is proposed [46] to enable data-driven agriculture.

These systems usually consist of low-powered sensors and are sparsely deployed in large areas with possibly intermittent connectivity. Therefore, sensors in the systems won't have careful maintenance, and the collected data can be noisy. Currently, IoT systems for agriculture lack a systematic method to evaluate the data quality of numerous datasets in this field. In this study, we propose a data-driven method to conduct data quality assessment without the requirement of farmers.

2.2 Sensor Calibration

Various approaches for sensor calibration are explored and can be categorized into two categories based on whether the reference truth is needed: non-blind calibration and blind calibration. For non-blind calibration, all methods rely on the accurate reference truth from expensive devices or other reliable sources to calibrate the sensors [48], such as linear regression [37] and maximum-likelihood [35]. Recently, machine-learning based methods are merged for reducing work effort for calibration [14, 33]. The blind calibration methods usually use geometric and physical constraints on the behavior [16] of sensors or oversample [7] to acquire additional information for calibration. Our work focus on assessing the sensor data quality, which is more informative for further information inference and/or system adaptation. These calibration approaches can not measure/assess how information of the sensor data.

2.3 IoT Data Quality

Multiple data quality assessment methods are proposed to determine the quality of acquired data. The concept of data quality varies for different projects and purposes. The prior works mostly focus on the aspect of validity, completeness, consistency, etc [10, 19, 27, 39, 42], which does not focus on the measurement of impacts to the knowledge inference accuracy. On the other hand, metrics have been explored to quantify the quality of the sensor signal, such as signal-to-noise ratio (SNR) [36] and structural similarity (SSIM) index [30]. These one-dimensional measurements are limited in the understanding of the causes of the low system performance. To resolve these challenges, schemes that measure multi-dimensional measurements of the data/signal quality are then explored for specific sensing modalities [21, 49]. For agriculture IoT applications, these data-driven approaches require certain reference truth or intensive measurements that are not scalable. A data-driven approach that does not require reference truth is needed.

3 *DAQUAL* DESIGN

Figure 2 shows the overview of *DaQual*, which contains three modules: contribution quantification module, sequence shuffle module, and data quality estimation module. First, a contribution quantification network (CQN) takes all sensor data as inputs and reports the fine-tuned parameters of the first layer as the contribution score, which indicates the contribution of one series to the prediction of another series (Section 3.1). Then, to make the different CQN calculated contribution score are comparable, the sequence shuffle module shuffles each sensor data as reference data and calculates a reference contribution score to normalize the contribution score (Section 3.2). Finally, the estimated contribution score and reference contribution score are then sent to Data Quality Estimation module to calculate the **normalized contribution score** and the data quality for the deployment in a farm (Section 3.3).

3.1 Contribution Quantification

We model the data quality assessment problem as a time series prediction problem between sensors deployed on the same farm, inspired by Granger causality [20, 43]. For two time series X_1 and X_2, if we utilize the history data of X_1 and X_2 together can help better predict the future data of X_2 than use X_2's historical data only, X_1 is said to be the 'Granger cause' of X_2 [38]. If X_1 is said to be the 'Granger cause' of X_2, there is an information flow between the X_1 and X_2 [9]. And the more information flows from X_1 to X_2, i.e., the more shared information, the more helpful of X_1 is in predicting X_2. Hence we leverage this principle to assist the estimation of

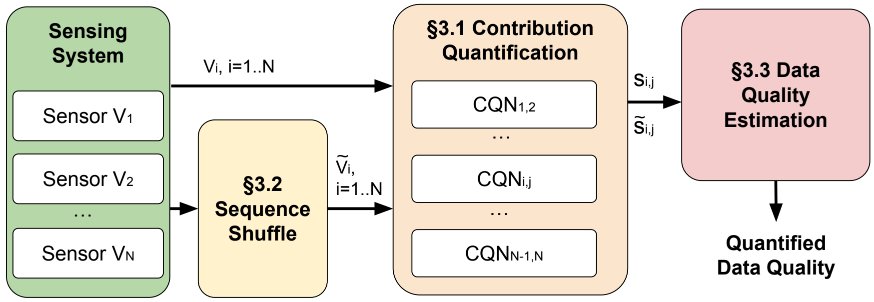

Figure 2: System overview. *DaQual* contains three modules: 1) contribution quantification, 2) sequence shuffle, and 3) data quality estimation. $V_0 \sim V_N$ are the sensor data, $\widetilde{V_i}$ is the shuffled sensor data, $s_{i,j}$ and $\widetilde{s_{i,j}}$ are the contribution score, and reference contribution score of sensor i on the prediction of sensor j, respectively.

data quality: the higher the quality, the more contribution from this sensor's information for the prediction.

We present the contribution quantification network (CQN) built upon the temporal convolutional network (TCN) architecture that infers causal relationships [34] between multiple time series. Figure 3 shows the architecture of CQN, which contains three parts: contribution score layer, TCN blocks, and the pointwise convolution layer. Given N sensor data, CQN first takes their data $V_1, ..., V_N$, as inputs. The contribution score layer initially selects one weight value for each input series and the output is the multiplier of the weight value and the input data. Then the outputs go through a TCN residual block to predict the values of each input series. Finally, the pointwise convolution layer integrates the prediction of each input series together as the prediction series $\widehat{V}$. In the training procedure, CQN calculates the loss between predicted series $\widehat{V_j}$ and the sensor data V_j, and updates the parameters of the network by the gradient descent algorithm [42]. The training of CQN will be terminated if the stop conditions are satisfied. CQN outputs trained parameters of the contribution score layer as the **contribution score** of sensor V_i on the prediction of sensor V_j. We note $s_{i,j}$ as the contribution score of sensor data $V_i, i = 1...N, i \neq j$ for the prediction of sensor data V_j. Different from the traditional neural network, the training purpose of CQN is to estimate the contribution score of each input series instead of a prediction model. Therefore, there is no training or testing set. For each input set, we train the network based on the historical data of all input series, and the outputs are the trained parameters of the contribution score layer.

3.1.1 Contribution Score Layer. The contribution score layer contains N nodes, and each node contains a weight parameter. All nodes are initialized with the same weight value, i.e., each time series has the same contribution to the prediction of the target series. To make the weight parameters trainable in the neural network, we utilize softmax function to convert the weight value of each node to a probability weight. The output of each node is the multiplier of the input series and the probability weight.

3.1.2 Temporal Convolutional Network (TCN) Blocks. The TCN blocks contain N TCN residual blocks and take the weighted series as inputs and output the prediction of each input series. As shown at the bottom of Figure 3, each TCN residual block contains two

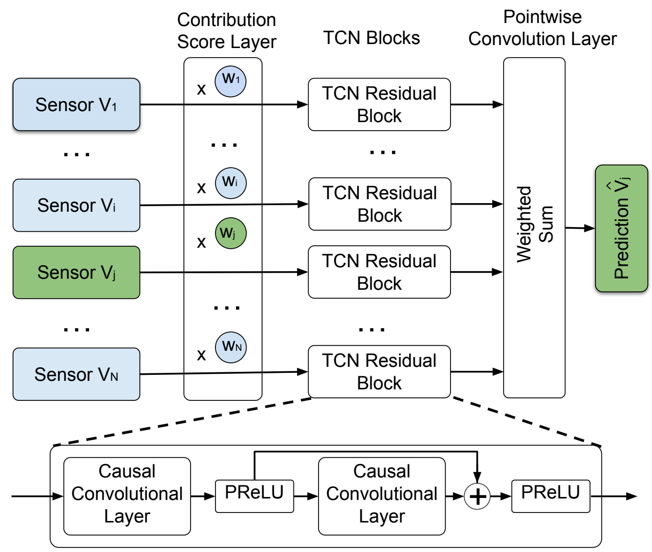

Figure 3: $CQN_{i,j}$ measures the contribution of sensor i to sensor j, $i = 1...N, i \neq j$. The network contains three parts: contribution score layer, TCN blocks, and pointwise convolution layer. The output of the CQN is the trained weight parameters of the contribution score layer.

causal convolutional layers [41] and two activation functions. A residual connection is added between the output of the first activation function and the input of the second activation function. TCN network employs a residual connection to encourage the model to learn the most representative features [25]. We adopt Parametric Rectified Linear Unit (PReLU) [24] as the activation function.

3.1.3 Pointwise Convolutional Layer. We further use a pointwise convolution layer [26] to integrate the prediction of N TCN blocks together as the output prediction $\widehat{V_j}$. The output of this layer is the weighted sum of the output of all TCN blocks. $\widehat{V_j}$ is the prediction series that has the same length as the input series (V_j).

3.1.4 Loss Function. We take one sensor data V_j as the ground truth to train the CQN, and adopt the mean square error (MSE) as the loss function to measure the difference between the ground truth (V_j) and the predicted series ($\widehat{V_j}$).

$$\text{MSE} = \frac{1}{\eta} \sum^{\eta} (V_j) - \widehat{V_j})^2$$

Where η is the length of the input series and the training goal is to minimize the MSE value.

3.2 Sequence Shuffle as Contribution Reference

The estimated contribution score $s_{i,j}$ of V_i in $CQN_{i,j}$ is not directly comparable to that of another $CQN_{i',j'}$, because their inputs are different. To make the contribution score of different sensors comparable, we further take the randomly shuffled sensor data $\widehat{V_i}$ as reference data to normalize the contribution score for each sensor. To normalize the contribution score $s_{i,j}$ of sensor data $V_i, i = 1...N, i \neq j$ for the prediction of sensor data V_j, *DaQual* calculates a reference contribution score with $\widetilde{V_i}$ that is shuffled. We consider the shuffled time sequence is a reference data that has no contribution to the prediction of sensor data V_j. Then we train another $CQN_{i,j}$ for the prediction of V_j, with $\widetilde{V_i}$ and other sensors data V_l, where $l = 1...N, l \neq i, l \neq j$. The output of this CQN is the reference contribution score $\widetilde{s}_{i,j}$. Therefore, we further calculate $A_{i,j} = s_{i,j} - \widetilde{s}_{i,j}$ and consider it as the normalized contribution sensor of V_i on the prediction of sensor V_j.

3.3 Data Quality Estimation

The data quality of a particular tree trunk sensor is then estimated as its contribution to the other sensors (of the same type) in the same farm/orchard. *DaQual* sums up the overall normalized contribution score $A_{i,j}(j = 1...N, j \neq i)$ for the sensor i data as its quantified data quality. The calculation of data quality of sensor i is as follows

$$Q_i = \sum_{j=1, j\neq i}^{N} A_i^j \tag{1}$$

We rank all sensors based on the quantified data quality Q_i of each sensor and select the top k sensors as the high-quality subset.

4 EVALUATION

4.1 Experiment Setup

Trunk Relative Water Content Dataset. We use the trunk relative water content dataset collected by sensors in 2018 [5], which includes nine sensors deployed on nine pomegranate trees. Figure 4 shows the details of the sensor locations. The data was collected between Aug. 22nd and Oct. 31st, 2018 per hour (71×24 samples per sensor). The stem water potential (SWP) data of each tree is measured by a pressure chamber [5] at noon on eight days of data collection. In this work, we considered the SWP value as the reference truth of the sensing data.

Evaluation Metric. We use the task of selecting k sensors of the highest data quality to evaluate the performance of the data quality assessment framework. We use the manually collected SWP data recorded around noon and from four directions of the tree (north, south, east, and west), as the **reference truth** Y. We then use samples of the sensor from 10 AM to 2 PM filtered with a

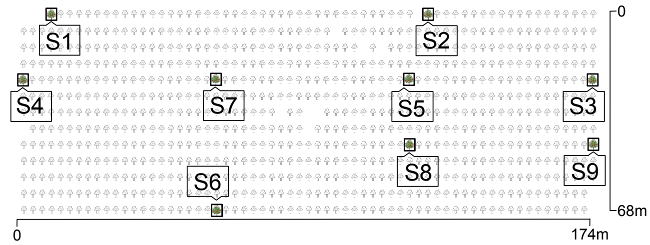

Figure 4: Sensor deployment locations. There are nine sensors (S1 to S9) deployed in different trees. Rows are spaced 5 meters apart and trees are spaced 2.75 meters apart in each row.

triangular window (centered 12 PM) [2] as the sensor data X. A linear regression model $\overline{Y} = a^* \cdot X + b^*$ is built between the k sensors' data X and their reference truth Y. Then we use R-square [] to quantify the performance of the selected k sensors. A higher R-square value indicates the selected subset of k has a better fitting with the reference truth, which we consider as better data quality. The R-square (R^2) is calculated as follows:

$$\begin{aligned} SST &= \sum_{i=1}^{h} (y_i - \bar{Y})^2 \\ SSR &= \sum_{i=1}^{h} \left(y_i - (a^* \cdot x_i + b^*)\right)^2 \\ R^2 &= 1 - \frac{SSR}{SST} \end{aligned} \tag{2}$$

Where h is the number of data points. In our experiment, there are $h = 8 \cdot k$ data points in total for the selected top k sensors over eight days that have reference truth.

Baseline Method. We evaluate *DaQual* against three baselines:

- Pearson correlation coefficient (PCC): the measurement of the correlation between two random variables [3]. We calculate PCC for each sensor pair. For sensor i, its quality is calculated as
$$\sum_j PCC_{i,j}, j = 1...N, j \neq i$$
The sensors with the highest top k quality are selected.
- Mutual information (MI): measures the mutual dependence between two variables [44]. We calculate MI for each sensor pair. For sensor i, its quality is calculated as
$$\sum_j MI_{i,j}, j = 1...N, j \neq i$$
The sensors with the highest top k quality are selected.
- Random selection. k sensors are randomly selected.

Implementation. To pre-process the data, we normalize each sensor's data to zero mean. We set the CQN kernel size (K) of the convolution filter as 15 and set the parameter of dilation and stride in *Conv1d* [11] both as 1. We use the Stochastic Gradient Descent algorithm and ADAM [28] as optimizers to train CQN. And we apply the early stopping method [12] to automatically stop

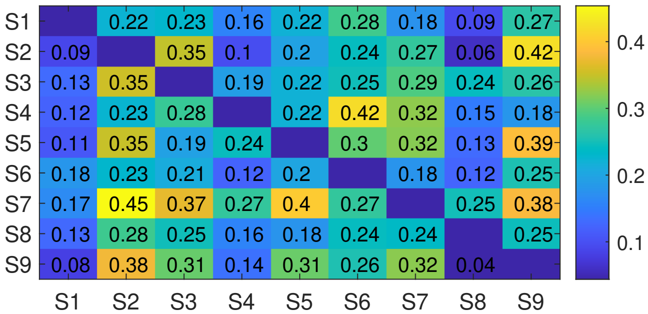

Figure 5: Normalized contribution matrix between sensors. Each element is the normalized contribution score of the column sensor to the prediction of the row sensor. For example, the normalized contribution score of sensor S9 to the prediction of sensor S2 is 0.42 (column 9, row 2). A higher normalized contribution score value means the column sensor has a more significant contribution to the prediction of the row sensor.

Table 1: R-square values of *DaQual* compared with the baselines for the selected top k high-quality subset.

k	*DaQual*	PCC	MI	Random
3	**0.937**	0.867	0.705	0.538
4	**0.933**	0.729	0.681	0.508
5	**0.837**	0.720	0.621	0.499

the training based on the loss decrease [47] with the factor and patience parameters set as 0.5 and 4, respectively. The training will be terminated when the learning rate drops to less than 0.001 (initially 0.01) or when the epoch achieves the maximum value of 6000. To decrease the bias induced by the randomization of shuffled sensor data and CQN initialization, we repeat the calculation 100 times and report the average value of the contribution score.

4.2 Results and Analysis

Figure 5 shows the normalized contribution score of *DaQual*. Each element is the normalized contribution score of the column sensor to the prediction of the row sensor. The quantified data quality Q_i of the sensor i is the sum of the ith column in the table.

4.2.1 R-square of Top k Sensors. We evaluate the top $k = 3, 4, 5$ sensor selection from the nine sensors. Table 1 shows the R-square of *DaQual* compared with the baselines from the selected top k high-quality sensors. *DaQual* achieves up to 0.4 (1.74×), 0.43 (1.84×), and 0.34 (1.68×) higher R-square values compared to the baselines for top 3, 4, and 5 subset, respectively. PCC and MI achieve a performance better than the random scheme. However, their decision is made based on the individual sensors' data and without a farm-level model to ensure the optimality of the subset.

4.2.2 Multi-Dimensional Scaling Based Distance Analysis. To better understand the present pattern, we utilize the multidimensional scaling tool [1] to visualize the relationships between sensors based on their quality. The multidimensional scaling function takes a symmetric distance matrix and outputs their relative positions. In our study, a higher normalized contribution score value $A_{i,j}$ indicates sensor i has a higher contribution to predicting j, which can be considered as a lower distance. Therefore, we calculate the distance $d(i, j)$ as

$$d(i, j) = \frac{T - A_{i,j} - A_{j,i}}{2}$$

Where $A_{i,j}$ is the normalized contribution score of sensing data j to sensing data i, and we set $T = 1$ empirically. Figure 6 shows the 2D scaling result of multi-dimensional scaling. We observe that the sensors with high-quality sensing data (#1 S2, #2 S9, #4 S3, #5 S7, #6 S5) are close to each other, and sensors with low-quality sensing data (#7 S4, #8 S8, #9 S1) are relatively far away from the high-quality sensing data.

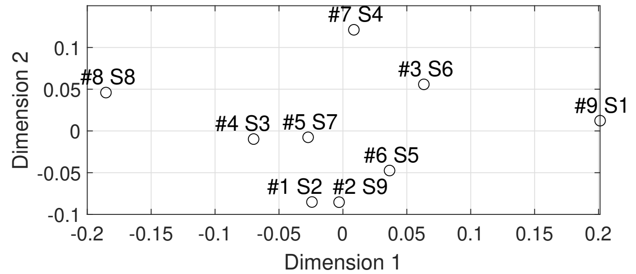

Figure 6: Multi-Dimensional Scaling based distance visualization. # ranks the data quality of each sensor.

5 CONCLUSION

In this paper, we present *DaQual*, a data-driven framework to assess the data quality of sensors on the farm. A temporal convolutional network-based scheme is presented to estimate the sensor quality as its contribution of the sensor in terms of predicting other sensors' time sequence data on the same farm. The neural network learns the underlying information across sensors in the same farm and the sensors' contribution is calculated from the parameters of the trained neural network. We conduct experiments on a real-world dataset of tree trunk relative water content sensors in a pomegranate orchard. The experiments show that *DaQual* achieves up to 1.8× improvement than the baselines for top k high-quality sensor selection.

ACKNOWLEDGEMENT

This material is based upon work primarily supported by the IoT4Ag Engineering Research Center funded by the National Science Foundation (NSF) under NSF Award Number EEC-1941529. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the author(s), and do not necessarily reflect those of the NSF.

REFERENCES

[1] 2022. Nonclassical multidimensional scaling. https://www.mathworks.com/help/stats/mdscale.html
[2] 2022. triangular window. https://www.mathworks.com/help/signal/ref/triang.html

[3] Jeremy Adler and Ingela Parmryd. 2010. Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry Part A* 77, 8 (2010), 733–742.

[4] Deepesh Agarwal, Pravesh Srivastava, Sergio Martin-del Campo, Balasubramaniam Natarajan, and Babji Srinivasan. 2021. Addressing Uncertainties within Active Learning for Industrial IoT. In *2021 IEEE 7th World Forum on Internet of Things (WF-IoT)*. 557–562. https://doi.org/10.1109/WF-IoT51360.2021.9595397

[5] A Alizadeh, A Toudeshki, R Ehsani, K Migliaccio, and D Wang. 2021. Detecting tree water stress using a trunk relative water content measurement sensor. *Smart Agricultural Technology* 1 (2021), 100003.

[6] Muhammad Ayaz, Mohammad Ammad-Uddin, Zubair Sharif, Ali Mansour, and El-Hadi M Aggoune. 2019. Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk. *IEEE access* 7 (2019), 129551–129583.

[7] Laura Balzano and Robert Nowak. 2007. Blind calibration of sensor networks. In *Proceedings of the 6th international conference on Information processing in sensor networks*. 79–88.

[8] Victor Blanco and Lee Kalcsits. 2021. Microtensiometers Accurately Measure Stem Water Potential in Woody Perennials. *Plants* 10, 12 (2021). https://www.mdpi.com/2223-7747/10/12/2780

[9] Katarzyna J Blinowska, Rafał Kuś, and Maciej Kamiński. 2004. Granger causality and information flow in multivariate processes. *Physical Review E* 70, 5 (2004), 050902.

[10] InduShobha N Chengalur-Smith, Donald P Ballou, and Harold L Pazer. 1999. The impact of data quality information on decision making: an exploratory analysis. *IEEE transactions on knowledge and data engineering* 11, 6 (1999), 853–864.

[11] PyTorch Contributors. 2022. CONV1D. https://pytorch.org/docs/stable/generated/torch.nn.Conv1d.html

[12] PyTorch Contributors. 2022. Reducelronplateau. https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html

[13] Ramide Augusto Sales Dantas, Milton Vasconcelos da Gama Neto, Ivan Dimitry Zyrianoff, and Carlos Alberto Kamienski. 2020. The swamp farmer app for iot-based smart water status monitoring and irrigation control. In *2020 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*. IEEE, 109–113.

[14] Saverio De Vito, Marco Piga, Luca Martinotto, and Girolamo Di Francia. 2009. CO, NO2 and NOx urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization. *Sensors and Actuators B: Chemical* 143, 1 (2009), 182–191.

[15] Muhammad Shoaib Farooq, Shamyla Riaz, Adnan Abid, Kamran Abid, and Muhammad Azhar Naeem. 2019. A Survey on the Role of IoT in Agriculture for the Implementation of Smart Farming. *Ieee Access* 7 (2019), 156237–156271.

[16] Jessica Feng, Seapahn Megerian, and Miodrag Potkonjak. 2003. Model-based calibration for sensor networks. In *SENSORS, 2003 IEEE*, Vol. 2. IEEE, 737–742.

[17] JE Fernández. 2014. Plant-based sensing to monitor water stress: Applicability to commercial orchards. *Agricultural water management* 142 (2014), 99–109.

[18] JE Fernández and MV Cuevas. 2010. Irrigation scheduling from stem diameter variations: a review. *Agricultural and Forest Meteorology* 150, 2 (2010), 135–151.

[19] Craig W Fisher, InduShobha Chengalur-Smith, and Donald P Ballou. 2003. The impact of experience and time on the use of data quality information in decision making. *Information systems research* 14, 2 (2003), 170–188.

[20] John R Freeman. 1983. Granger causality and the times series analysis of political relationships. *American Journal of Political Science* (1983), 327–358.

[21] Ruiyang Gao, Wenwei Li, Yaxiong Xie, Enze Yi, Leye Wang, Dan Wu, and Daqing Zhang. 2022. Towards Robust Gesture Recognition by Characterizing the Sensing Quality of WiFi Signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–26.

[22] Laura García, Lorena Parra, Jose M Jimenez, Jaime Lloret, and Pascal Lorenz. 2020. IoT-based smart irrigation systems: An overview on the recent trends on sensors and IoT systems for irrigation in precision agriculture. *Sensors* 20, 4 (2020), 1042.

[23] Syeda Iqra Hassan, Muhammad Mansoor Alam, Usman Illahi, Mohammed A Al Ghamdi, Sultan H Almotiri, and Mazliham Mohd Su'ud. 2021. A systematic review on monitoring and advanced control strategies in smart agriculture. *IEEE Access* 9 (2021), 32517–32548.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*. 1026–1034.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[26] Binh-Son Hua, Minh-Khoi Tran, and Sai-Kit Yeung. 2018. Pointwise convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 984–993.

[27] Aimad Karkouch, Hajar Mousannif, Hassan Al Moatassime, and Thomas Noel. 2016. Data quality in internet of things: A state-of-the-art survey. *Journal of Network and Computer Applications* 73 (2016), 57–81.

[28] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[29] Ankur Kohli, Rohit Kohli, Bhupendra Singh, and Jasjit Singh. 2020. Smart plant monitoring system using IoT technology. In *Handbook of Research on the Internet of Things Applications in Robotics and Automation*. IGI Global, 318–366.

[30] Chaofeng Li and Alan C Bovik. 2009. Three-component weighted structural similarity index. In *Image quality and system performance VI*, Vol. 7242. SPIE, 252–260.

[31] Leslie Lipper, Philip Thornton, Bruce M Campbell, Tobias Baedeker, Ademola Braimoh, Martin Bwalya, Patrick Caron, Andrea Cattaneo, Dennis Garrity, Kevin Henry, et al. 2014. Climate-smart agriculture for food security. *Nature climate change* 4, 12 (2014), 1068–1072.

[32] Yuehua Liu, Tharam Dillon, Wenjin Yu, Wenny Rahayu, and Fahed Mostafa. 2020. Noise removal in the presence of significant anomalies for industrial IoT sensor data in manufacturing. *IEEE Internet of Things Journal* 7, 8 (2020), 7084–7096.

[33] Balz Maag, Olga Saukh, David Hasenfratz, and Lothar Thiele. 2016. Pre-Deployment Testing, Augmentation and Calibration of Cross-Sensitive Sensors.. In *EWSN*. 169–180.

[34] Meike Nauta, Doina Bucur, and Christin Seifert. 2019. Causal discovery with attention-based convolutional neural networks. *Machine Learning and Knowledge Extraction* 1, 1 (2019), 312–340.

[35] Boon Chong Ng and Chong Meng Samson See. 1996. Sensor-array calibration using a maximum-likelihood approach. *IEEE Transactions on Antennas and Propagation* 44, 6 (1996), 827–835.

[36] Alan V Oppenheim and Ronald W Schafer. 1975. Digital signal processing(Book). *Research supported by the Massachusetts Institute of Technology, Bell Telephone Laboratories, and Guggenheim Foundation. Englewood Cliffs, N. J., Prentice-Hall, Inc., 1975. 598 p* (1975).

[37] Antonios E Panteleon, Kerstin Rebrin, and Garry M Steil. 2003. The role of the independent variable to glucose sensor calibration. *Diabetes Technology & Therapeutics* 5, 3 (2003), 401–410.

[38] David A Pierce and Larry D Haugh. 1977. Causality in temporal systems: Characterization and a survey. *Journal of econometrics* 5, 3 (1977), 265–293.

[39] Leo L Pipino, Yang W Lee, and Richard Y Wang. 2002. Data quality assessment. *Commun. ACM* 45, 4 (2002), 211–218.

[40] S Pooja, DV Uday, UB Nagesh, and Shamitha Gurunath Talekar. 2017. Application of MQTT protocol for real time weather monitoring and precision farming. In *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*. IEEE, 1–6.

[41] Josh Robinson, Scott Kuzdeba, James Stankowicz, and Joseph M Carmack. 2020. Dilated causal convolutional model for RF fingerprinting. In *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 0157–0162.

[42] Sebastian Ruder. 2016. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747* (2016).

[43] Anil Seth. 2007. Granger causality. *Scholarpedia* 2, 7 (2007), 1667.

[44] Ralf Steuer, Jürgen Kurths, Carsten O Daub, Janko Weise, and Joachim Selbig. 2002. The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics* 18, suppl_2 (2002), S231–S240.

[45] Sumya Tabassum, Afzal Hossain, et al. 2018. Design and development of weather monitoring and controlling system for a smart agro (farm). *Intelligent Control and Automation* 9, 03 (2018), 65.

[46] Deepak Vasisht, Zerina Kapetanovic, Jongho Won, Xinxin Jin, Ranveer Chandra, Sudipta Sinha, Ashish Kapoor, Madhusudhan Sudarshan, and Sean Stratman. 2017. {FarmBeats}: An {IoT} Platform for {Data-Driven} Agriculture. In *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17)*. 515–529.

[47] Yuan Yao, Lorenzo Rosasco, and Andrea Caponnetto. 2007. On early stopping in gradient descent learning. *Constructive Approximation* 26, 2 (2007), 289–315.

[48] Lei Zhang, Fengchun Tian, Chaibou Kadri, Bo Xiao, Hongjuan Li, Lina Pan, and Hongwei Zhou. 2011. On-line sensor calibration transfer among electronic nose instruments for monitoring volatile organic chemicals in indoor air quality. *Sensors and Actuators B: Chemical* 160, 1 (2011), 899–909.

[49] Yue Zhang, Zhizhang Hu, Susu Xu, and Shijia Pan. 2021. AutoQual: task-oriented structural vibration sensing quality assessment leveraging co-located mobile sensing context. *CCF Transactions on Pervasive Computing and Interaction* 3, 4 (2021), 378–396.

[50] Yang-Yang Zheng, Jian-Lei Kong, Xue-Bo Jin, Xiao-Yi Wang, Ting-Li Su, and Min Zuo. 2019. CropDeep: the crop vision dataset for deep-learning-based classification and detection in precision agriculture. *Sensors* 19, 5 (2019), 1058.