accepted to appear in Proceedings of the 46th IEEE Conference on Local Computer Networks (LCN), October 4-7, 2021.

Towards Dense and Scalable Soil Sensing Through Low-Cost WiFi Sensing Networks

Steven M. Hernandez, Deniz Erdag and Eyuphan Bulut
Department of Computer Science, Virginia Commonwealth University
401 West Main St. Richmond, VA 23284, USA
{hernandezsm, erdagd, ebulut}@vcu.edu

***Abstract*—Precision agriculture uses precise sensor data collected throughout farmland to give farmers better insight into their land, allowing for greater crop yields and reduced resource usage. However, existing solutions require high hardware costs thus limiting large scale deployments. To address that, we propose a low-cost and scalable solution for sensing physical attributes of soil using IoT based WiFi sensing devices. By understanding variations in WiFi radio signals with channel state information (CSI) and machine learning models, we evaluate the proposed soil sensing system through experiments on physical soil traits such as soil moisture content, soil texture and position. Moreover, we also demonstrate how a mesh network of WiFi sensing devices allows us to predict the physical traits of the soil in the area between each pair of sensors, allowing for an increase in sensing area coverage as nodes are added.**

***Index Terms*—WiFi sensing, sensor networks, soil sensing, soil moisture, precision agriculture.**

I. Introduction

Growing global population paired with environmental changes requires the development of innovative solutions to ensure sufficient future food production. Projections propose that by 2050, agricultural food production will need to increase upwards of 70% [1]. Precision agriculture aims to increase commercial farmland yields while using available resources (e.g., water) efficiently through sensor data collection and advanced prediction models [2]–[4]. One group of important metrics for farmland sensing are the physical soil attributes such as the moisture content and texture, where the latter is usually described by the mixture of silt, clay, sand, and stones. While the soil texture is generally static over time, some attributes of the soil are time variant. For example, soil moisture content will vary depending on the time since the last rain fall or last manual irrigation. Additionally, these attributes will vary based on unique soil texture properties and based on the time of day as moisture evaporates from the soil.

A key issue in data collection for precision agriculture is the sparseness of sensor deployment. *Direct methods* for measuring soil moisture are considered the most accurate option [5] because they directly measure water content, while *indirect methods* measure some related property such as resistance or capacitance. Even so, direct methods such as the gravimetric method require soil samples to be physically removed from the environment resulting in destructive and non-continuous sensing of the moisture level. Furthermore, this method requires a 24 hour oven drying time for every measurement. Alternatively, there are some sensors that can sense the soil moisture continuously and provide measurements over time. However, they can only sense data in the immediate region where the sensor is placed. Moreover, such sensors are mostly expensive (e.g., Tensiometers ranging from $\$100$ to upwards of $\$1,000$ [6] per sensor) thus their dense deployment across large farmlands is not affordable by most farmers. While some low-cost sensors exist for sensing soil moisture, they are mostly unreliable and do not provide commercial-grade accurate measurements, thus are only used by hobbyist.

Radio frequency-based sensing methods have appeared as alternatives for tracking soil attributes such as moisture content [7]. However, most of the signal-based systems (e.g., ground penetrating radar (GPR) [7], [8]) require high cost transmitting and receiving equipment. A more recent work addresses this by considering commodity WiFi devices [9] as a lower cost solution for wireless soil sensing. Even so, the cost and size of the required hardware is still quite significant and as such, only a single receiver device was deployed within the soil. With our recently developed ESP32 CSI Toolkit [10], we can collect channel state information (CSI) of WiFi signals from a very low-cost ($< \$5$), standalone IoT microcontroller. This allows for wider deployments of WiFi sensing devices enabling soil sensing within a network of these devices. Soil properties can then be tracked between each pair of WiFi sensing transceiver devices. Furthermore, the WiFi network can be used to communicate this sensed data without requiring additional hardware to a centrally located database for further processing and analysis thus further reducing the costs associated with deploying our proposed system.

In this work, we explore the use of our proposed WiFi sensing system for tracking soil attributes using machine learning. We evaluate the use of CSI for predicting soil properties such as soil moisture levels, soil texture and positioning. Finally, we evaluate how our proposed sensing system architecture shown in Fig. 1c can distinguish soil properties with greater coverage compared to standard sensor-based approaches through the use of a mesh WiFi sensing network.

The rest of the paper is organized as follows. In Section II, we discuss background information about precision agriculture and WiFi sensing. In Section III, we discuss our proposed system including details about the hardware and machine learning architectures as well as information about our experimental methods using this system. We continue with the

analysis and evaluation of the experiments in Section IV. Then we show how IoT WiFi soil sensing nodes can be used to develop a mesh network with exponentially growing coverage in Section V. Lastly, we conclude this work with our final remarks in Section VI.

II. Background

Precision agriculture (PA) aims to gather large amounts of analytical data about farmland through the use of IoT sensing devices distributed throughout the farmland. After processing this data, the goal is to give farmers more insight into temporal characteristics of their farms as well as spatially localized suggestions such as in precision irrigation decision making [2]. PA covers all aspects of farmland management through sensors and advanced information technology techniques such as field tracking through UAVs [4] as well as modeling spatial and temporal crop yield predictions through machine learning [11]. Wireless sensor networks (WSN) are typically deployed [3] to aggregate the data from the distributed sensors to a single central location such as a cloud storage for further processing using cloud computation platforms.

WiFi sensing uses the signal data collected from WiFi devices such as laptops, phones and routers to understand physical attributes of the environment through signal variations caused by Line-of-Sight (LoS) obstructions, Non-LOS multi-path interference as well as phase shifts caused by time delay. Using WiFi for soil sensing in PA is useful because it can reduce the hardware cost while also increasing sensing coverage because the area between each pair of transceiver devices is sensed. Furthermore, WiFi can be used for dual purposes, first for wireless sensing of soil traits and second as a communication method in a WSN. One common metric used for WiFi sensing is channel state information (CSI) [10]. CSI is used in orthogonal frequency division multiplexing (OFDM) systems to estimate signal propagation information details over multiple subcarrier frequencies. CSI ($\mathbf{H}$) is given as an $M \times N$ matrix where M is the number of antennas and N is the number of subcarriers. To identify the value for $\mathbf{H}$, we consider the equation $y = \mathbf{H}x + \eta$ where x is the transmitted signal, y is the received signal and η is some noise from the environment. Given the real component (h_r) and imaginary component (h_{im}) for each subcarrier i within H, we can calculate the amplitude (A) and phase (ϕ) by:

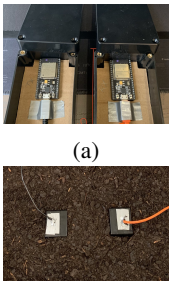
$$A = \sqrt{(h_r)^2 + (h_{im})^2}, \tag{1}$$

$$\phi = atan2\left(h_{im}, h_r\right). \tag{2}$$

For our experiments, we use a single antenna for both the transmitter and receiver using standard 2.4GHz channels with a bandwidth of 20MHz which allows us to collect values for 64 subcarriers where 52 are non-null. Signal propagation for frequency f at a distance d can be modeled as

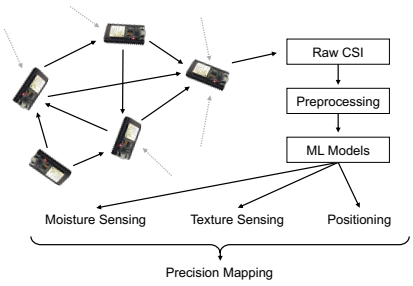
$$E(f, d) = Ae^{-(\alpha + j\beta)d}, \tag{3}$$

where α is an attenuation coefficient due to the physical environment between the transmitter and receiver resulting from soil content, and β is a phase coefficient caused by propagation in the soil environment. Both α and β can be modeled as functions of permittivity (a complex value $\epsilon^* = \epsilon' + j\epsilon''$) and electrical conductivity (σ) by:



(a)

(b)



(c)

Fig. 1: (a) ESP32 transmitter (TX) and receiver (RX) in waterproof container. (b) TX and RX placed in soil-filled container for an experiment. (c) Architecture diagram for our proposed WiFi sensing mesh network for soil sensing.

$$\alpha = \frac{2\pi f}{c}\sqrt{\frac{\epsilon'_r}{2}[\sqrt{1 + tan^2\delta} - 1]}, \tag{4}$$

$$\beta = \frac{2\pi f}{c}\sqrt{\frac{\epsilon'_r}{2}[\sqrt{1 + tan^2\delta} + 1]}, \tag{5}$$

where the speed of light $c \approx 3 \times 10^8 m/s$, and the loss tangent

$$tan\delta = \frac{\epsilon''_r + \frac{\sigma}{2\pi f \epsilon_0}}{\epsilon'_r} \tag{6}$$

as described in [9]. Permittivity and electrical conductivity metrics have long been used for understanding many soil attributes as discussed in [12].

III. Proposed Soil Sensing System

In this work, we propose a low-cost system to understand physical attributes of soil using standard 2.4 GHz WiFi signals. To accomplish this, we need a tool to collect CSI data for both training and evaluating our machine learning models as well as for deploying our system in a network in the field. Our proposed system architecture together with the devices used is shown in Fig. 1.

CSI Collection Tool: To collect CSI, we use ESP32 WiFi-enabled microcontrollers and run our ESP32 CSI Toolkit[1]. This low-cost microcontroller can be used for initial data collection experiments and can be deployed into full networks as a standalone device. During our initial experiments, the CSI receiver is connected directly to an Android smartphone running our custom CSI annotation application to record and annotate all CSI data and associated metadata for future analysis. In large scale deployments, the ESP32 devices can use their WiFi hardware not only for WiFi soil sensing but also for communicating results between neighbors and to edge devices for data aggregation and precision mapping [4].

[1]https://stevenmhernandez.github.io/ESP32-CSI-Tool/

Machine Learning Architecture: To make our predictions, we select a Dense Neural Network (DNN) classifier architecture with two hidden dense layers with identical number of hidden neurons, each followed by a dropout layer used to prevent overfitting by setting the output of neurons randomly to 0 with a probability of $P_{dropout} = 0.2$. We use a softmax activation function and Stochastic Gradient Descent (SGD) to optimize our loss function

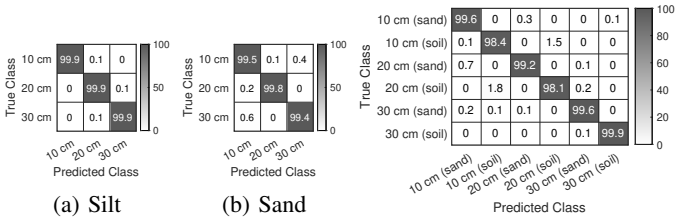
$$\mathcal{L}(x, y) = \frac{1}{N}\sum_{i=1}^{N}(\mathcal{F}(x_i) - y_i)^2, \tag{7}$$

where $\mathcal{F}$ is the model, x_i is the i-th input sample for the model and y_i is the label for the i-th sample. We use learning rate $\eta \in \{0.001, 0.01, 0.1\}$ to control the speed at which the model converges during training. As input to the classifier we give a matrix of size $w \times n$, where w is a window of CSI measurements and $n = 64$ is the number of subcarriers. Our preprocessing and machine learning steps are written in Python using the Keras [13] deep learning library. Hyperparameter optimization was used to balance model accuracy with computation time. For reproducibility in our experiments, we use the Sacred experiment database Python library [14] to capture results for each experiment along with the chosen hyperparameter values.

Experiment Process: To evaluate the proposed system for soil sensing tasks, we use our ESP32 toolkit to collect and annotate CSI to predict the following physical soil traits: depth in soil, distance in soil between the transmitter (TX) and receiver (RX), soil moisture level, and soil texture. For each experiment, we use two ESP32s, one TX and one RX which are both housed inside waterproof enclosures as shown in Fig. 1a. For soil depth and TX/RX distance experiments, the enclosures are placed directly in the soil for each position as shown in Fig. 1b. For soil texture and soil moisture experiments, we constructed a jig to hold the ESP32s stationary across multiple repetitions. We collect CSI for each state in an experiment using a timer to ensure a similar number of samples are collected per class. Annotated data for each experiment is split such that the first 50% of samples per class are used for training and the final 50% of samples are used for testing.

IV. Experimental Results

To explore the use of WiFi for soil sensing, we evaluate our CSI collection system and machine learning prediction models for the following tasks. First, we identify if the model is able to recognize different depths and distances between transmitter and receiver in different soil types in Section IV-A. This first set of experiments is useful to ensure that the model is able to detect the physical location even when devices are placed underground. Next, we look at how well our model is able to distinguish moisture levels in Section IV-B. Soil moisture is a dynamic metric which changes over time and must be maintained through precision irrigation to ensure proper crop growth. Following this, we also look at how soil texture can be detected with our model in Section IV-C. Soil texture is defined by the distribution of clay, sand and silt



(a) Silt (b) Sand (c) Silt and Sand

Fig. 2: Confusion matrix for soil depth sensing.

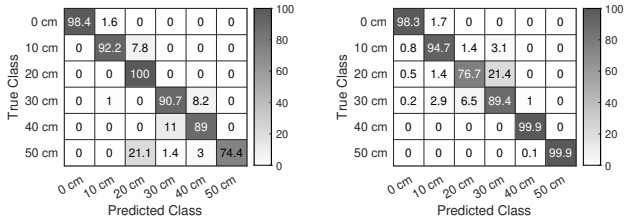
content and is an important factor which affects temporal soil moisture properties. Next, we perform two null hypothesis experiments in Section IV-D to ensure that our model classification results are valid and not due to other spurious physical attributes within the environment. The importance of these null hypothesis experiments stems from our goal to achieve generalizable results for our soil sensing experiments. Finally, we also provide long-term moisture sensing results with the proposed system in Section IV-E.

A. Detecting TX Depth and TX-RX Distance

When WiFi sensing devices are deployed in real world scenarios, each device will be subject to unique physical placement throughout the farmland. As such, we want to see how well the models will be able to judge depth within the soil as well as distance between transmitter and receiver. Additionally, because each sensor location will be unique, the soil texture will also be diverse. As such, we want to make sure that our model is not confused or tricked by the presence of different soil types. To check this, we experiment with two distinct soil types: pure *sand* and pure *silt*.

First, we consider the question of recognizing depth in soil. We record three depths 10 cm, 20 cm and 30 cm in both sand and silt environments. The RX is placed above the soil while the TX is placed in the soil at the given depths. In Fig. 2a we can see that the prediction accuracy for each depth is 99.9% for the silt environment while the sand environment in Fig. 2b achieves an accuracy of 99.5%, 99.8% and 99.4% across each respective depth. This demonstrates that WiFi CSI can be used to easily distinguish the depth of the node within the soil. To further evaluate these experiments, we consider the case where our model should distinguish both depth and soil type. Fig. 2c shows that the model can achieve accuracy above 98.1% across each of these six classes.

Next, we look at the performance of the model in predicting distances between TX and RX. Both TX and RX are placed in the soil at distances $d \in \{0, 10, 20, 30, 40, 50\}$ cm (with depth 5 cm). For the silt environment, we achieve an overall accuracy of 90.77% while the sand environment achieves an overall accuracy of 93.15%. For the silt environment, the confusion matrix in Fig. 3a shows that the true class 50 cm achieves the lowest prediction accuracy of 74.4% while each of the other distances achieves greater than 89.0%. The sand environment on the other hand reaches a low accuracy of 76.7% for the 20 cm class while each of the other classes in the experiment are able to achieve greater than 89.4% as shown in Fig. 3b.



(a) Silt (b) Sand

Fig. 3: Confusion matrix for soil distance sensing.

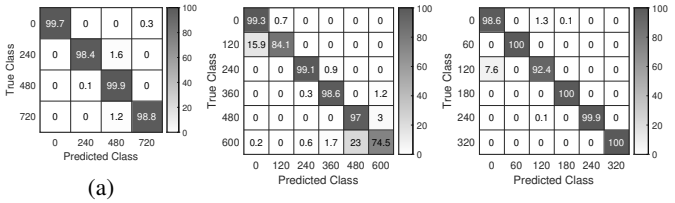
$d^{(silt)}$ \ $d^{(sand)}$	0cm	10cm	20cm	30cm	40cm	50cm
0cm	88.7	96.2	100.0	99.8	99.9	100.0
10cm	98.6	99.9	100.0	99.9	100.0	99.8
20cm	99.1	99.8	100.0	99.9	100.0	100.0
30cm	99.4	100.0	100.0	99.8	99.9	99.9
40cm	99.9	100.0	100.0	99.7	99.4	100.0
50cm	100.0	97.3	98.8	99.8	99.5	98.4

TABLE I: Binary classification accuracy as distance changes for sand versus silt when both TX and RX are under the soil.

In order to ensure that the model is able to distinguish between the two environments of silt versus sand, we also compare the binary classification accuracy for different distances $d^{(silt)}$ for distance in the silt environment and $d^{(sand)}$ for distance in the sand environment. When $d^{(silt)} = d^{(sand)} = 0$ cm, we find the lowest accuracy of 88.7%. This is because there is no soil between the TX and RX, thus the direct LOS signal is not directly affected by the soil content. We might expect that the model would achieve only 50% binary classification accuracy as a result, however, because our CSI metric is affected by the multipath environment, some received signal is still able to bounce into the soil environment which then bounces back to the receiver antenna. Thus, the model is able to still distinguish the two soil types. As the values for $d^{(silt)}$ and $d^{(sand)}$ vary, the binary classification accuracy achieves greater than 96.2% for each other position. Looking down the diagonal of the table, we can see that similar distances for the silt and sand environments are distinguishable from one another with an accuracy of 98.4% and greater other than when $d^{(silt)} = d^{(sand)} = 0$. This shows that similarly positioned nodes in different soil mediums can distinguish the soil medium best when there is soil in the LOS, but NLOS can also achieve relatively high accuracy through the received multipath signals. Overall the accuracy is 99.26% on average for these binary classifiers.

B. Detecting Soil Moisture

Soil moisture is an important dynamic metric to track over time for general agricultural tasks. Direct methods for collecting water content in soil such as the gravimetric method [5] measures the weight of a soil sample taken from the environment before and after a 24 hour oven drying sequence. The difference in weights before and after drying corresponds directly to the amount of water lost during the drying step.



(a) (b) (c)

Fig. 4: Moisture sensing accuracy when using different moisture level increments. (a) $240x$ mL increments (overall accuracy: 99.19%), (b) $120x$ mL increments (overall accuracy: 92.09%), (c) $60x$ mL increments (overall accuracy: 98.49%).

For our experiments, we begin by drying the soil for 3 hours in an oven set to $100°$ Celsius to ensure a consistent baseline moisture content. Then, for each unique moisture level, we mix 1200 mL of dry soil with v mL of water. CSI is collected for 10 seconds per container in consecutive order over 10 repetitions. For the water content, we begin with large distinct level increments of 240 mL (i.e., around a cup) where $v = 0$ mL is the default dry soil and $v = 720$ mL is our highest moisture level. The confusion matrix for this first experiment given in Fig. 4a shows that the accuracy for each class ranges from 98.4% to 99.9%, and the overall accuracy for the model on all four classes together is 99.19%.

As a result of this high predictive accuracy, we reduce the moisture level increments down to 120 mL which achieves an overall accuracy of 92.09% in Fig. 4b and again reduce to 60 mL increments in Fig. 4c which gives an overall accuracy of 98.49%. The experiment with 120 mL increments achieves the lowest accuracy of 92.09% compared to the other two soil moisture experiments. From the confusion matrix, we can see that this is the result of relatively low predictive accuracy on two classes: 120 mL and 600 mL. For the 120 mL class, we can see that the accuracy for the class is 84.1% where 15.9% of class samples are incorrectly predicted as the 0 mL class while for 600 mL, the accuracy is lower at 74.5% where 23.0% of samples are predicted to be class 480 mL. From this, we can see that the misclassification for these classes is typically found in the adjacent moisture levels showing that there is some overlap in the data distributions of nearby moisture levels. The relatively lower performance for 120 mL can also be found in Fig. 4c which shows a lower predictive accuracy (92.4%) relative to other classes in the same 60 mL increments experiment due to the confusion with 0 mL class. Since the error is found in both cases at 120 mL, this could be a result of slight inaccuracies in that specific soil moisture sample. Collecting more data with more samples can potentially ensure a better accuracy from the model, however we leave this for our future work.

C. Detecting Soil Texture

Soil texture descriptions are composed of the percentage of clay separate, silt separate and sand separate found within a given sample of soil. Variations in soil texture for an environment directly affect the hydraulic properties of the soil

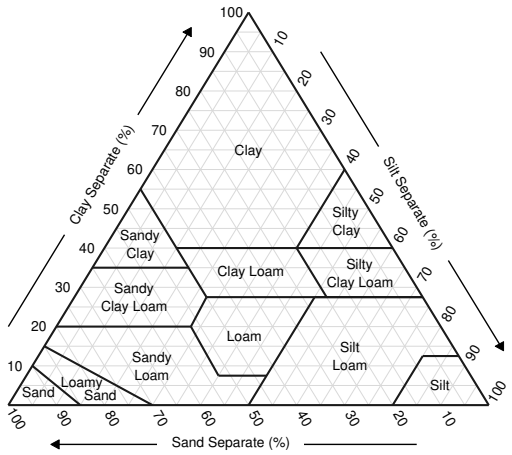


Fig. 5: Soil texture triangle.

c_1 / c_2	Clay	Silty Clay	Silty Clay Lm.	Silt Lm.
Silt	100.00%	99.99%	100.00%	97.60%
Silt Lm.	73.53%	65.58%	85.60%	
Silty Clay Lm.	94.95%	79.06%		
Silty Clay	54.64%			

TABLE II: Binary classification accuracy (Clay to Silt).

in the environment [15] and thus have an important impact on soil moisture. The soil texture triangle in Fig. 5 is commonly used to label different types of soil. To evaluate our system on soil texture recognition, we create individual soil samples representing different classes found in the soil texture. As a large number of classes are present in the soil texture triangle, we look at the binary classification accuracy between pairs of texture classes to identify how well the system is able to distinguish the classes.

We begin by evaluating the accuracy of our system on five classes along the *"Silt Separate"* axis on the soil triangle. Specifically, we create a model which predicts from the following five classes: *Clay*, *Silty Clay*, *Silty Clay Loam*, *Silt Loam* and *Silt*. The results in Table II show that our model can easily distinguish the two furthest classes (*clay* and *silt*). In fact, if we look at the *silt* row from the table we can see that silt can be distinguished from every other class with an accuracy greater than or equal to 97.60%. Clay on the other hand can be confused with *Silt Loam* (i.e., 73.53% accuracy) and *Silty Clay* (i.e., 54.64% accuracy) indicating that the binary classification model is unable to detect any distinguishing features between the two classes. On average, the binary classification models are able to achieve an accuracy of 85.10%.

We continue evaluation of texture classes from the *"Sand Separate"* axis of the soil triangle, namely classes *Silt*, *Silt Loam*, *Loam*, *Sandy Loam* and *Sand*[2]. The results for these models are shown in Table III. Similar to the previous results, the furthest classes *Silt* and *Clay* are able to be classified by the model with an accuracy of 99.98%. Also similarly, if we consider the *silt* column, we can see that *silt* can easily be

[2]Loamy sand is not considered as we find it very similar to sand.

c_1 / c_2	Silt	Silt Lm.	Loam	Sandy Lm.
Sand	99.98%	95.60%	76.61%	80.00%
Sandy Lm.	99.79%	73.20%	99.90%	
Loam	100.00%	86.56%		
Silt Lm.	97.60%			

TABLE III: Binary classification accuracy (Silt to Sand).

c_1 / c_2	Sand	Sandy Lm.	San. Cl. Lm.	Sandy Clay
Clay	76.95%	64.35%	60.53%	63.95%
Sandy Clay	60.28%	91.72%	99.17%	
San. Cl. Lm.	99.58%	99.75%		
Sandy Lm.	80.00%			

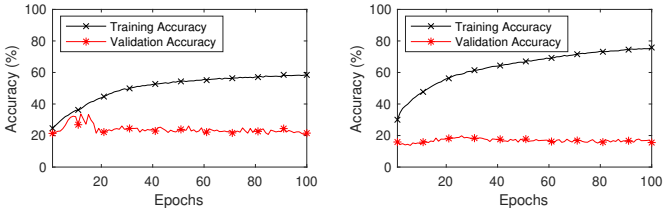
TABLE IV: Binary classification accuracy (Sand to Clay).

distinguished between all other classes with a minimum of 97.60% accuracy. On average, the binary classification models for these classes are able to achieve 90.93% accuracy.

Finally, we consider the *"Clay Separate"* axis with classes *Sand*, *Sandy Loam*, *Sandy Clay Loam*, *Sandy Clay* and *Clay* with results shown in Table IV. We find that *sand* and *clay* are not able to achieve as high accuracy as the other two pairs of furthest classes (i.e., only 76.95% accuracy). The overall binary classification accuracy for all class pairs is also low (i.e., 79.63%) compared to previous axis classes. Our initial insight for these lower results is due to the similarity of visual appearance and feel of the *sand* and *clay* soils. However, we will look into this further in our future work.

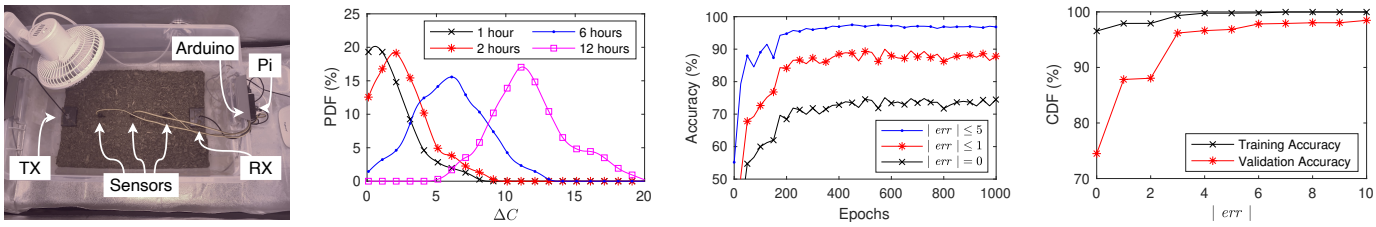
D. Null Hypothesis

To ensure that our model is predicting the unique soil traits that we specified rather than some other temporal traits in the environment, we take two precautions. First, as mentioned previously, we repeat all sets of classes over multiple repetitions for each experiment. This ensures that temporal changes in the signal are not what is being detected by the model and also adds variety to the samples in both the training and testing phases. Second, we perform two additional experiments here to check the null hypothesis. We follow the same setup as the previous experiments but this time in the first null hypothesis experiment, each container is empty and in the second experiment, each container is filled with the same homogeneous mixture of soil. Similar to the moisture experiments we perform, we use 6 containers for both experiments. For the empty container experiment,



(a) (b)

Fig. 6: Null hypothesis results. (a) Empty containers. (b) Filled containers.



(a) (b) (c) (d)

Fig. 7: (a) Experiment setup for long-term automated moisture monitoring system. (b) Probability density function (PDF) for the change in capacitance reading (C) over $h \in \{1, 2, 6, 12\}$ hour time-spans. (c) Accuracy during training when allowable error $|err| \leq e \in \{0, 1, 5\}$. (d) Training and validation accuracy as allowable error ($|err|$) increases.

as it is shown in Fig. 6a, the accuracy of the model on the training set achieves 58.46% while the accuracy on the unseen validation set is much lower at 21.54%. Similarly, the experiment with filled containers achieves 75.83% accuracy on the seen training set while unseen validation set achieves only 15.54% accuracy. With 6 classes, we would expect a randomly predicting model that achieves around 16.67% accuracy, which is approximately what we are seeing in the validation accuracy for both experiments. On the other hand, the training accuracy is able to achieve higher accuracy because the model is trained directly on this data which allows the model to memorize the data rather than learn distinct generalizable traits of the data distribution. This shows that in the previous experiments the model is in fact learning the physical attributes from the soil.

E. Long-Term Moisture Sensing

To evaluate our system for more realistic conditions, we perform a long running experiment using our system where moisture is allowed to naturally evaporate over time. To accomplish this, we begin by placing one TX, one RX and three capacitive moisture sensors into a large bucket of gardening soil. Since the soil remained indoors for the duration of the experiment, a dehumidifier and a fan are used to increase the speed of drying. Three capacitive moisture sensors are connected to an Arduino Pro Micro which is then connected through USB serial along with the ESP32 TX and RX into a Raspberry Pi 4 Model B running our custom data logger software. A photo of our system is shown in Fig. 7a. For this experiment, CSI and moisture values are recorded for 30 second segments every 1 hour over a timespan of 5 days. At the beginning of the experiment, water is mixed thoroughly into the soil to ensure a homogeneous moisture level throughout. Over the experiment timespan, the capacitance value (C) increases from 300 up to 350. In Fig. 7b we look at the distribution of the change in capacitance (ΔC) over different periods of time. This shows that over most of the one-hour segments, C increased 0 or 1 units with an average of approximately $\Delta C = 0.6$. Subsequently, over all of the 2 hour segments, the capacitance increased by 2 and similarly for the 6 hour segments, we see a peak at $\Delta C = 6.1$. For the 12 hour segments, the peak shows that most of the 12 hour segments that we recorded show an increase of capacitance $\Delta C = 11.1$. Interestingly though, we can see a long tail to these distributions showing that, for example, some 12 hour segments can result in an increase close to $\Delta C = 20$.

Soil Property	Accuracy	Figure	#Reps.	Time/Rep.
Depth (Silt)	99.91%	Fig. 2a	3	30 sec.
Depth (Sand)	99.58%	Fig. 2b	3	30 sec.
Depth (Silt/Sand)	99.13%	Fig. 2c	3	30 sec.
Distance (Silt)	90.77%	Fig. 3a	6	30 sec.
Distance (Sand)	93.15%	Fig. 3b	6	30 sec.
Distance (Silt/Sand)	99.26%	Table I	12	30 sec.
Moisture (240 mL)	99.19%	Fig. 4a	10	10 sec.
Moisture (120 mL)	92.09%	Fig. 4b	10	10 sec.
Moisture (60 mL)	98.49%	Fig. 4c	10	10 sec.
Texture (clay/silt)	85.10%	Table II	10	10 sec.
Texture (silt/sand)	90.03%	Table III	10	10 sec.
Texture (sand/clay)	79.63%	Table IV	10	10 sec.
Null Hyp. (empty)	23.82%	Fig. 6a	10	10 sec.
Null Hyp. (filled)	15.54%	Fig. 6b	10	10 sec.
Moisture (Long)	74–97%	Fig. 7c	1	5 days

TABLE V: Overview of performed soil sensing experiments.

We train our model over $1,000$ epochs on our training dataset and show the accuracy for our validation dataset over these epochs in Fig. 7c. After training, the model achieves an accuracy of 74.48% for predicting the capacitance value between $300 \leq C \leq 350$. As we noted in Fig. 7b, a small change in capacitance (i.e., $\Delta C = 1$) is only related to approximately 1 hour of drying time. As such, we can allow our model some additional margin of error when making predictions. The red line in Fig. 7c shows the case when we allow a margin of error $|err| \leq 1$ where $err = \hat{y} - y$ such that $\hat{y}$ is the predicted capacitance value and y is the true recorded capacitance value. After training fully, when $|err| \leq 1$ we achieve an accuracy of 87.83% while we achieve an even higher accuracy of 96.84% when $|err| \leq 5$. In Fig.7d, we show the final accuracy after training for different margin of errors ($|err|$). We observe that increasing $|err|$ from 0 to 1 offers an increase of $+13.35\%$. Similarly, increasing $|err|$ from 2 to 3 offers an increase of $+8.13\%$ going from 88.06% up to 96.19%. Beyond this, increasing from $|err| = 3$ up to 10 only offers an increase of $+2.29\%$. Thus, by allowing a margin of error of just $\Delta C = \pm 3$ we achieve a high accuracy of 96.19%.

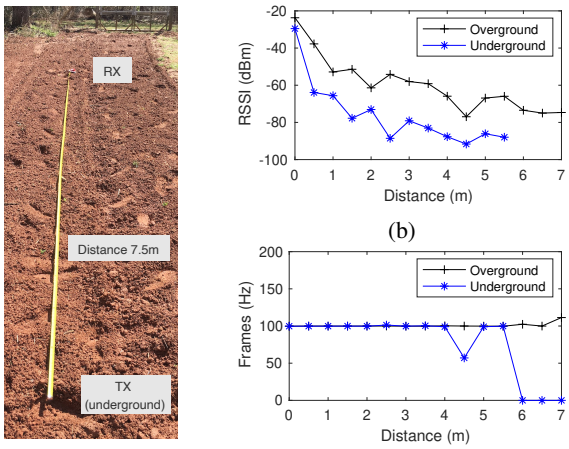
Results Summary. Table V shows a summary of the experimental results in this section together with the settings used.

V. Scalable Mesh Network

Thanks to the low cost of the ESP32 microcontrollers used in our proposed system, a large number of devices can be deployed to cover large areas of farmland. Through a mesh network, sensed data can be routed to an aggregation location for further processing without requiring each device to be directly connected to a router. This is important because in large farmland deployments, it is not expected that WiFi signals of standard routers will cover the entire area. Instead, each ESP32 can perform both WiFi sensing and WiFi routing to share information. Moreover, each ESP32 can also act as both transmitter and CSI receivers. As a result of this, each ESP32 can then capture raw CSI from multiple neighboring transmitters. Preprocessing this raw CSI and inputting it into pretrained machine learning models allows us to perform moisture sensing, texture sensing and positioning tasks which can then be used together to calculate precision mapping [4].

A. Communication within Soil

When radios are placed into soil, signal attenuation increases compared to the case where devices are placed in open air line-of-sight positions. As such, we must consider the maximum distance that devices are able to communicate before the signal is overrun by environmental noise. We perform this experiment with a depth of 0 cm (overground) and a depth of 25 cm (underground) in an outside garden as shown in Fig. 8a. In Fig. 8b we plot the Received Signal Strength Indicator (RSSI) for each distance. RSSI is a more course-grained signal metric than CSI, but gives us a single metric value for each received frame and follows the log-distance path model [16] such that $d = 10^{\frac{A-RSSI}{10\times n}}$, where d is the predicted distance, A is a baseline RSSI measurement at 1 meter, $RSSI$ is the current RSSI measurement and n is the path-loss exponent for the given environment. From Fig. 8b, we can see that the underground value has a lower RSSI for all distances compared to the overground setting. This is because the signal is more attenuated in the underground case. A more important metric in particular for WiFi soil sensing is the maximum distance until WiFi CSI frames can no longer be received. In the overground case, frames can be received beyond 7 m, while in the underground case, no frames can be received at or beyond 6 m. The lowest average RSSI value for the underground scenario is RSSI $= -91.64$ dBm at $d = 4.5$ m. This lowest RSSI value also corresponds to a decrease in received CSI frames as shown in Fig. 8c. The transmitter sends frames at a consistent 100Hz but at $d = 4.5$ m, the receiving rate decreases to an average of 57Hz. This is likely because the signal strength of the frames are dropping below the noise floor which makes it harder for the receiver to successfully decode the transmitted frame thus resulting in dropped frames. Interestingly, with $d \in \{5.0, 5.5\}$, the CSI frame rate returns back to 100Hz before dropping down to 0Hz at $d \geq 6.0$. This return to normal corresponds to the slight rise in average RSSI back above -90 dBm. We can also see that the lowest average RSSI (-76.94 dBm) for the overground scenario is recorded at $d = 4.5$ m, as in the underground scenario. Similarly we can see some relationship between the peaks and valleys in both experiments. These variations occur as a result of the multipath environment as well as Fresnel Zones [10] which can have both constructive and destructive impacts on the signal amplitude.



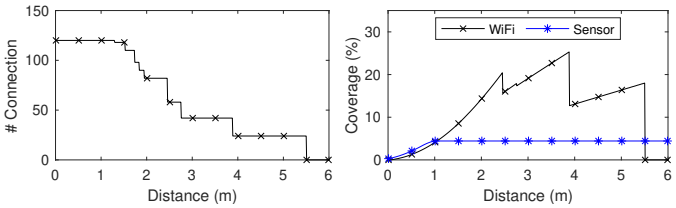
(a) (b) (c)

Fig. 8: Signal attenuation is greater within soil than through air. (a) Photo of the experiment taking place in a garden with tilled soil. (b) RSSI measurements between TX and RX, when TX is placed overground (depth $= 0$ cm) and underground (depth $= 25$ cm). (c) CSI frames collected per second at different distances.

B. Scalability of Wireless Sensing

Using this knowledge, we propose the use of a mesh network of WiFi soil sensing nodes. In typical sensor deployments, each additional sensor increases the sensing coverage linearly, however with our mesh network, because each device can act as both TX and RX, each added sensor can increase the coverage exponentially. We consider 16 sensor locations within an 18 m $\times$ 18 m area where the sensors are laid out in a 4×4 grid with a distance of d m between neighboring nodes. For our simulation, we design this area as a bitmap where each 100×100 pixel section is equivalent to a one meter square area. For the WiFi sensing evaluation, each location corresponds to an ESP32 transceiver node which can act as both TX and RX. Thus, each device within communication distance $d_{comm.} = 5.5$ m can perform soil sensing in the areas between the two devices. To model the coverage between pairs of nodes, an ellipse is placed such that the major vertices correspond to the position of the TX node (P_{TX}) and RX node (P_{RX}) with a width $w = 50$ cm on the minor axis and a height $h = \|P_{TX} - P_{RX}\|$, the euclidean distance between them.

In Fig. 9a, we can see that as the grid distance d increases between neighboring nodes, the number of pairwise connections decreases. When $0 \leq d \leq 1.29$ m, all devices are within communication range of each other and thus WiFi sensing can be performed between each pair. With 16 nodes, there are 120 unique connections possible. When $d = 0$, the number of connections is very high, but the sensed area for each of



(a) (b)

Fig. 9: Results of mesh grid coverage simulation as the distance (d) between neighboring nodes increases. (a) Number of connections between pairs of WiFi devices. (b) Percent of area covered by WiFi-sensing and sensor-based method.

these device pairs contains many overlapping regions; thus, a very low coverage is achieved as shown in Fig. 9b. After $d = 1.29$ m, the number of connections begins to decrease, however we do not see a decrease in coverage until $d = 2.46$ m as only overlapping sensing areas are lost. After the sudden decrease at $d = 2.46$ m an even higher peak is reached at $d = 3.88$ m with 25.3% area coverage even though the number of connections is reduced to only 42 connections. This is because with smaller d, there are more overlapping sensing areas while with larger d, each pair of devices covers more distinct areas. After $d = 3.88$ m, another dip occurs followed by a rise up until $d = 5.5$ m where WiFi signals are no longer able to be reliably received in the soil. Thus, after $d = 5.5$ m, the coverage drops immediately to 0%.

Unlike our proposed soil sensing mesh network, standard sensor-based approaches use multiple sensors which collect readings in small localized area. Thus, to compare with sensor-based systems, we use the same grid and set the sensing diameter to 1 m around each node. Note that we allow this diameter to be larger than the 50 cm width used for the ellipse-based coverage in our WiFi sensing simulation. Fig. 9b shows that the coverage of the sensor approach begins slightly higher than the WiFi approach before quickly plateauing at a coverage percentage of 4.4% after $d = 1$ m at which point coverage neither increases nor decreases because of the localized sensing provided by each sensor. On the other hand, the sensor-based approach can still achieve some coverage (4.4%) even after $d = 5.5$ m but it is far lower than the highest coverage value that can be obtained with the proposed WiFi sensing approach (25.3%).

VI. Conclusion

Future precision agriculture techniques will allow for better crop growth tracking and thus increase crop yield globally. Through our proposed system, the placement of WiFi sensing devices can increase the sensing coverage density while reducing the number of deployed sensors. This allows for high precision mapping of sensor data across farmland while also reducing the cost of deployment. We do this by leveraging low-cost WiFi-enabled microcontrollers to reduce the costs for deploying the system and simplifying the deployment. In this work, we demonstrate how our proposed WiFi soil sensing system can be used for sensing soil moisture, soil texture and position within the soil. Furthermore, we demonstrate how our system can be used in long-term deployments to track moisture levels over time. In a real-world field we evaluate the WiFi transmission capability of our system underground to understand limits of the signal communication range. Finally, we show how the use of WiFi soil sensing allows for greater coverage of a given area by leveraging multiple neighboring transceivers compared to standard sensor deployments which can only sample sensor data in very localized areas.

Acknowledgements

This work is supported in part by National Science Foundation (NSF) Graduate Research Fellowship Program (GRFP) under Grant No. 1744624, Virginia Commonwealth University Presidential Research Quest Fund (PeRQ) and Commonwealth Cyber Initiative (CCI).

References

[1] M. C. Hunter, R. G. Smith, M. E. Schipanski, L. W. Atwood, and D. A. Mortensen, "Agriculture in 2050: recalibrating targets for sustainable intensification," *Bioscience*, vol. 67, no. 4, pp. 386–391, 2017.

[2] C. Dewi and R.-C. Chen, "Decision making based on iot data collection for precision agriculture," in *Asian Conference on Intelligent Information and Database Systems*. Springer, 2019, pp. 31–42.

[3] R. Priya and D. Ramesh, "Ml based sustainable precision agriculture: a future generation perspective," *Sustainable Computing: Informatics and Systems*, p. 100439, 2020.

[4] D. Vasisht, Z. Kapetanovic, J. Won, X. Jin, R. Chandra, S. Sinha, A. Kapoor, M. Sudarshan, and S. Stratman, "Farmbeats: An IoT platform for data-driven agriculture," in *14th {USENIX} Symp. on Networked Systems Design and Implementation ({NSDI}*, 2017, pp. 515–529.

[5] V. Sharma, *Methods and techniques for soil moisture monitoring*. University of Wyoming Extension, 2018.

[6] R. Prince, "Soil moisture monitoring: a selection guide," 2019. [Online]. Available: https://www.agric.wa.gov.au/horticulture/soil-moisture-monitoring-selection-guide

[7] A. Salam and U. Raza, *Signals in the Soil: An Introduction to Wireless Underground Communications*. Cham: Springer International Publishing, 2020, pp. 3–38. [Online]. Available: https://doi.org/10.1007/978-3-030-50861-6_1

[8] L. Zhou, D. Yu, Z. Wang, and X. Wang, "Soil water content estimation using high-frequency ground penetrating radar," *Water*, vol. 11, no. 5, p. 1036, 2019.

[9] J. Ding and R. Chandra, "Towards low cost soil sensing using Wi-Fi," in *The 25th Annual International Conference on Mobile Computing and Networking*, 2019, pp. 1–16.

[10] S. M. Hernandez and E. Bulut, "Lightweight and standalone IoT based WiFi sensing for active repositioning and mobility," in *21st International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM) (WoWMoM 2020)*, Cork, Ireland, Jun. 2020.

[11] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Computers and electronics in agriculture*, vol. 151, pp. 61–69, 2018.

[12] A. Brovelli and G. Cassiani, "Combined estimation of effective electrical conductivity and permittivity for soil monitoring," *Water Resources Research*, vol. 47, no. 8, 2011.

[13] A. Gulli and S. Pal, *Deep learning with Keras*. Packt Publishing Ltd, 2017.

[14] K. Greff, A. Klein, M. Chovanec, F. Hutter, and J. Schmidhuber, "The sacred infrastructure for computational research," in *Proceedings of the 16th Python in Science Conference*, vol. 28, 2017, pp. 49–56.

[15] S. Bousbih, M. Zribi, C. Pelletier, A. Gorrab, Z. Lili-Chabaane, N. Baghdadi, N. Ben Aissa, and B. Mougenot, "Soil texture estimation using radar and optical data from Sentinel-1 and Sentinel-2," *Remote Sensing*, vol. 11, no. 13, p. 1520, 2019.

[16] S. M. Hernandez and E. Bulut, "TrinaryMC: Monte Carlo based anchorless relative positioning for indoor positioning," in *IEEE 17th Annual Consumer Com. & Networking Conference (CCNC)*, 2020, pp. 1–6.