SoilCares: Towards Low-cost Soil Macronutrient and Moisture Monitoring Using RF-VNIR Sensing

Juexing Wang¹*, Yuda Feng²*, Gouree Kumbhar³, Guangjing Wang¹, Qiben Yan¹, Qingxu Jin¹, Robert C. Ferrier¹, Jie Xiong², Tianxing Li¹
¹Michigan State University, East Lansing, Michigan, USA
²University of Massachusetts Amherst, Amherst, Massachusetts, USA
³The Dow Chemical Company, Deer Park, Texas, USA

{wangjuex@msu.edu,wudafeng@umass.edu,∗, kumbharg,wanggu22,qyan}@msu.edu,{billjin,ferrier5}@egr.msu.edu,jxiong@cs.umass.edu,litianx2@msu.edu

ABSTRACT
Accurate measurements of soil macronutrients (i.e., nitrogen, phosphorus, and potassium) and moisture play a key role in smart agriculture. However, existing commodity soil sensors are often expensive and the achieved accuracy is unsatisfactory. To address these issues, we present SoilCares, a low-cost soil sensing system enabling accurate and simultaneous monitoring of the concentration levels of soil moisture and macronutrients. SoilCares overcomes key challenges of accommodating diverse soil types and soil textures by introducing a novel membrane-based scheme. For moisture sensing, SoilCares leverages the multi-modal fusion of RF and NIR signals to significantly increase the sensing accuracy. Through delicate hardware design, we enable negligible-cost sensor data transmission using the existing sensing hardware, building up a complete end-to-end soil sensing system. SoilCares is cost-effective ($63.5), portable (0.5 kg), and low-power (236 μW), making it suitable for in-situ deployment. On-site experimental results show that SoilCares achieves high macronutrient sensing accuracy with a low RMSE of 0.138, and extremely low moisture estimation error of 1%, outperforming the state-of-the-art research and expensive commodity moisture sensors on the market.

CCS CONCEPTS
• Human-centered computing → Ubiquitous and mobile computing systems and tools.

KEYWORDS
Smart agriculture, Pervasive sensing, Multi-modality, Soil sensing Low-power and low-cost sensing

ACM Reference Format:

1 INTRODUCTION
Precision agriculture, which refers to precise water, nitrogen, phosphorus, and potassium control in different farm regions, depends on accurate soil moisture and macronutrient sensing. Proper fertilization enhances grain yields, while overusing fertilizers can lead to pollution of aquatic systems [74] and groundwater [15]. Moreover, proper soil moisture level facilitates plant absorption of essential nutrients, and precision irrigation has been proposed to save precious water resources and enable sustainable agriculture [24, 27, 33, 39]. Therefore, the ability to monitor the concentration levels of macronutrients and moisture has become a vital component in smart agriculture.

In recent years, several low-cost soil moisture sensing systems have been proposed, and they adopt various RF signals, including Wi-Fi, LoRa, LTE, and RFID techniques [13, 20, 22, 70]. However, these systems still lack the critical fertilizer sensing capability. The fundamental drawback that prohibits them from accurate fertilizer sensing is that RF waves with centimeter-level or even millimeter-level wavelengths are too coarse-grained to detect the variation of nutrients. In comparison, Vis-NIR (visible–near-infrared) waves have a nanometer wavelength, which is at the same scale as nutrient molecules and manifests unique physical properties of absorption/reflectance. As a result, Vis-NIR (VNIR) reflectance spectroscopy has become a prevalent method for analyzing soil properties [50, 55, 59]. However, this method requires a high-end spectrometer (~$20,000 [8]) for generating a super-high-resolution reflectance spectrum and extracting nutrient-relevant information, which also introduces significant maintenance overhead. Moreover, it requires intricate pre-processing, such as grinding and drying, to mitigate the impacts of soil moisture, particle size, and other environmental factors. This has hindered the wide deployment of on-site fertilizer sensing.

To address the issues of reflectance spectroscopy while still leveraging the advantages of VNIR sensing, low-cost LEDs and photodiodes (PDs) are emerging as viable alternatives in agriculture sensing [5]. Combinations of LEDs can cover the discrete spectrum range from VIS to NIR, and photodiodes can serve as receivers of the reflected light. These devices are cheap, portable, and adaptable to various working environments with fewer restrictions. However, there are three major challenges associated with low-cost soil sensing systems. First, pre-processing steps like drying and grinding are critical for soil sensing due to significant influences of soil moisture and particle size on estimating other soil properties [49–51, 59, 72]. This is because soil moisture significantly affects the accuracy of macronutrient sensing. We need to totally remove it (i.e., drying) or estimate it accurately to remove its effect on macronutrient sensing. However, current low-cost RF-based solutions [13, 20, 22, 70] are...
Without the grinding step, the large and random soil particle size mainly contributes of SoilCares are as follows:

SoilCares is the first system demonstrating the capability of low-error of 1% for moisture estimation. To the best of our knowledge, concentration measurements, respectively, and an extremely low mean-square error (RMSE) of 0.144, 0.147, and 0.125 for N, P, and K in real-world experiments show that SoilCares achieves the root-mean-square error of 0.175, 0.175, and 0.138 for macronutrient sensing under varying conditions including different soil types/mixtures and various moisture levels. The proposed soil moisture sensing module achieves 1% mean absolute error, outperforming the state-of-the-art research [13, 20, 22] and expensive commodity moisture sensors on the market.

2 BACKGROUND

2.1 NIR-based Soil Macronutrient Sensing

The Beer-Lambert law [46] describes the attenuation of light intensity as it traverses through a substance, linking it to the material’s constituent. This law is frequently utilized in chemical analysis to evaluate the concentration of chemical components capable of light absorption and scattering [48]. As the light of a specific spectrum passes through substances such as N, P, and K, it provokes the molecular bonds of each component to vibrate. Due to its unique molecular structure and bond, every chemical species generates a distinctive absorption spectrum, which can be used for element analysis [26, 64]. The absorbance of substance A can be derived in as below [65]:

\[ A = \log \left( \frac{I_0}{I_r} \right) = \gamma \ell c, \]  

where \( I_0 \) is the emitted light intensity, \( I_r \) is the received light intensity after propagating through the optical length \( \ell \), \( \gamma \) is the molar attenuation coefficient, and \( c \) is the concentration of the attenuating species.

2.2 RF-based Soil Moisture Sensing

Volumetric water content (VWC) \( \theta_W \), the volume of water per unit volume of soil, is a common metric for soil moisture measurement. Prior studies [34, 62, 67] have revealed the dependence of the dielectric constant \( \epsilon \) on VWC. The empirical formula to quantify the relationship between \( \epsilon \) and \( \theta_W \) [67] is depicted below:

\[ \theta_W = 0.1138 \sqrt[1758]{\epsilon} - 0.1758. \]  

Based on Eq. 2, RF-based soil moisture sensing has been proposed [20]. By measuring the RF wave propagation speed \( c \) in the target soil, the dielectric constant can be obtained as \( \sqrt[1758]{\epsilon} = c_0/c \), where \( c_0 \) denotes the RF signal propagation speed in the air. So, soil

![Figure 1: SoilCares system for macronutrients (N, P, K) and soil moisture measurements, with negligible-cost communication capability](image-url)
moisture can be estimated by measuring the RF signal propagation speed in the soil.

**TDoF-based soil moisture sensing.** Conventional RF solutions such as Time Domain Reflectometry (TDR) radar [38] estimate the RF wave speed based on accurate Time-of-Fight (ToF) measurements, which, however, require ultra-wide bandwidth, escalating the hardware cost. To enable affordable solutions, Time-Difference-of-Flight (TDoF) based approaches have been proposed [13, 20, 22].

As shown in Fig. 2, the same RF signal is transmitted by two antennas through soil and air and eventually received at the receiver. Since the distance between transmitter and receiver is much larger than twice the wavelength (e.g. 33 cm for LoRa at 915 MHz), the propagation paths can be considered as parallel [53]. The TDoF of the two paths can then be obtained as:

$$\Delta t = \frac{\Delta d_3 + \Delta d_2}{c} - \frac{\Delta d_1}{c_0} = \frac{\Delta d_3 + \Delta d_2 - \Delta d_1}{c}. \quad (3)$$

According to Snell’s law [73] and the geometric relationship, we have:

$$\sin \theta_1 = \frac{\lambda_1}{c} = \frac{c}{V}, \quad \sin \theta_2 = \frac{\lambda_2}{c} = \frac{c}{V}$$

where $\lambda_1$ and $\lambda_2$ are the RF wavelength in the air and in the soil, respectively.

With Eq. 4, the term $\Delta d_2 - \Delta d_1 / V$ in Eq. 3 is canceled out, and the relationship between dielectric constant $\varepsilon$ and TDoF $\Delta t$ can be obtained as:

$$\varepsilon = \left(\frac{c_0}{c_0 \Delta t} \right)^2 = \left(\frac{c_0}{\Delta d_1 \cos \theta_2} \right)^2. \quad (5)$$

In Eq. 5, $\Delta d$ denotes the spacing between two antennas, which is predefined. The incident angle $\theta_2$ does not need to be known and can be approximated with a constant value [13]. Therefore, the soil moisture $\theta_2$ could be estimated from the TDoF $\Delta t$ given Eq. 2 and 5. TDoF $\Delta t$ can be accurately calculated by measuring the phase difference $\Delta \phi$ of the received signals transmitted from the two antennas [20, 22]:

$$\Delta t = \frac{\Delta \phi}{2\pi f_c}. \quad (6)$$

where $f_c$ denotes the carrier frequency of the RF signal.

### 3 SOILCARES DESIGN

In this section, we first describe the overview of SoilCares illustrated in Fig. 3, followed by the four design components. The first component involves techniques to eliminate the impact of soil moisture and particle size in macronutrient monitoring, allowing SoilCares to operate outside laboratory scenarios. Second, we present multi-modal soil moisture monitoring, combining RF and NIR sensing. With the multi-modal fusion, we enhance the accuracy and reliability of soil moisture sensing and further improve the sensing accuracy of macronutrients by eliminating the significant effect of moisture. Third, we address the monitoring of soil nitrogen, phosphorus, and potassium by designing a cost-effective LED-PD array. Last, negligible-cost LoRa transmission is realized on the existing sensing hardware, facilitating real-world deployments.

### 3.1 Elimination of Confounding Factors for Soil Macronutrients Sensing

Recent studies have demonstrated the potential of macronutrient monitoring using VNIR optical systems, thanks to the rich soil spectral information contained within this spectrum [51, 55, 64]. Most of these studies, however, are predominantly confined to laboratory environments due to the requirement of pre-processing steps such as drying and grinding. These steps are crucial for mitigating the influence of soil moisture and particle size on the reflectance spectroscopy of soil samples [51, 55]. Soil moisture impacts the absorption spectra of soil elements, and different particle sizes lead to inconsistent reflectance, resulting in varying levels of accuracy [49–51, 59, 72].

We introduce a module to eliminate noise from soil moisture and particle size. To tackle the issue of varying soil particle sizes, we employ a membrane as depicted in Fig. 10(g). The membrane acts as an exchange platform, absorbing moisture and macronutrients until the equilibrium state is reached, i.e., the moisture and macronutrient levels are the same in the membrane and in the soil. Compared to directly measuring the moisture and macronutrient levels in the soil, the membrane presents a surface with more evenly distributed moisture and macronutrients, allowing for more accurate and stable measurements. Instead of selecting irregular soil particles as the surface for reflectance, this membrane absorbs macronutrients and water from the soil and provides a uniform surface for VNIR light reflectance measurements. Since the membrane is made of Poly-Vinylidene-Fluoride-co-Hexafluoropropylene (PVDF-co-HFP), it has minimal impact on Radio Frequency (RF) signals. We describe the membrane manufacturing details in Sec. 4.
To accommodate the noise caused by soil moisture, we break down the spectral absorbance $A_i$ of the soil sample at the $i^{th}$ wavelength into three subcomponents as illustrated in Eq. (7):

$$A_i = A_{ti} + A_{mi} + A_{oi},$$

where $A_{ti}$ is the spectral absorbance of the target macronutrient, $A_{mi}$ is the absorbance of soil moisture content, and $A_{oi}$ is the absorbance of other elements. Given that the water component significantly influences the absorption spectrum at 1450 nm due to the O-H bond’s vibration [7, 11, 41, 44], we establish a mapping function $\phi()$. This function links the absorbance value at 1450 nm, represented as $A_{m1450}$, to each $A_{mi}$ at various soil moisture levels.

By eliminating the mapped component $\phi(A_{m1450})$ from the total absorbance $A_i$ at the $i^{th}$ wavelength, we obtain the remaining absorbance $A_{ri}$ in Eq. 8:

$$A_{ri} = A_i - \phi(A_{m1450}) = A_{ti} + A_{oi}. \quad (8)$$

Subsequently, the concentration values $y$ and the absorbance $A_{ri}$ are projected into a latent space using Least Squares Support Vector Regression (LS-SVR) as:

$$Y = \sum_{j=1}^{k} \alpha_j f(A_{ri}) + b_2,$$

where $Y$ represents the projected concentration values $y$ in the latent space, $f()$ is the linear function that projects the value of $A_{ri}$ into a new latent space, $k$ is the total count of new latent variables based on the original number of wavelengths $n$, $\alpha_j$ represents the loading vector for the latent variable $f(A_{ri})$, and $b_2$ is the bias of the new variable in the latent space. The initial absorbance of other elements, denoted as $A_{oi}$, exhibits inconsistent trends across different concentration levels, and its magnitude is substantially lower than the absorbance of the target macronutrient, $A_{ti}$. The impact of unrelated variables is further diminished through the projection process, resulting in a reduced amplitude. Consequently, we can disregard the value of $A_{oi}$ in Eq. 7 and approximate the value of $A_{ri}$ to Eq. 9.

### 3.2 NIR-RF-based Soil Moisture Sensing

While both NIR and RF measurements can be used for soil moisture estimation, they are prone to errors caused by environmental factors. Specifically, NIR measurements can be affected by soil granularity and porosity [10, 64]. RF-based soil sensing is biased by uneven surfaces and obstacles such as stones [13, 22]. Interestingly, the factors affecting NIR and RF sensing are orthogonal, which facilitates fusing NIR and RF sensing for better performance.

The learning-based sensing techniques [32, 61, 68, 69, 71] have been widely studied. We explore the characteristics of both NIR and RF modalities and propose a multi-modal moisture sensing model. Specifically, we use the raw measurements, i.e., signal phase from RF receivers, and NIR reflectance from photodiode as input features to train the machine learning model. The obtained model is then used to predict soil moisture, which efficiently mitigates the bias from each individual modality and achieves higher accuracy than either of them.

For NIR-based moisture sensing, we use an LED transmitting 1450 nm light with a photodiode receiver since water molecules majorly absorb 1450 nm light waves [47]. NIR reflectance manifests a monotonic empirical relationship [41, 42] with the soil moisture level. For RF sensing, we use a LoRa transmitter with a single-input dual-out RF switch to enable two-antenna transmission. By switching channels within a chirp, we see a sudden phase jump due to the propagation path difference between two antennas, thus enabling TDoF sensing at the receiver side. The RF dielectric permittivity change due to different soil moisture levels is nonlinear based on Eq. 2. Consequently, the fusion task is modeled as a monotonic regression problem with two independent features.

To minimize the latency, we develop a fusion model based on small-size machine learning models instead of deep neural networks. Specifically, we compare ten machine learning models, including six linear and nonlinear models, three ensemble learning methods, and a dimensionality reduction method. The architecture of both 1-layer and 2-layer neural networks are determined by grid search on our collected dataset. The training dataset comprises around 600 soil samples, including three soil types across 5-50% soil moisture levels. We used 6-fold cross-validation in the training. The 2-layer neural network has two inputs, i.e., NIR reflectance and RF signal phase, ten neurons for each of the two hidden layers along with ReLU activation, and one output of the soil moisture estimation. The 1-layer neural network has the same inputs/output and 50 neurons in the hidden layer. Adam Solver is used for both neural networks, with a constant learning rate of 0.001.

Table 1 shows the mean absolute error of soil moisture level estimation across three soil types (see Sec. 5.2 for detailed comparison) and their combinations. The Decision Tree and Random Forest methods outperform others. Decision Tree achieves the highest accuracy on a single soil type but performs not as good as Random Forest considering all soil types. It indicates that Decision Tree is more prone to overfitting. Thus, we choose the Random Forest model for multi-modal fusion, and it achieves much higher accuracy (an error of 0.13%) than the high-end commodity soil sensor [3] (an error of 2.54%). Sec. 5.3.1 will present more evaluation on other practical considerations.

### 3.3 Spectral Channel Selection for Soil Macronutrient Sensing

Having addressed the interference of soil moisture and particle size on soil macronutrient sensing, the subsequent challenge we face

<table>
<thead>
<tr>
<th>Categories</th>
<th>ML Method</th>
<th>Soil 1</th>
<th>Soil 2</th>
<th>Soil 3</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Ridge</td>
<td>1.37</td>
<td>1.33</td>
<td>2.83</td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td>LS-SVR</td>
<td>1.21</td>
<td>1.33</td>
<td>2.49</td>
<td>1.80</td>
</tr>
<tr>
<td>Basic nonlinear</td>
<td>Polynomial</td>
<td>1.37</td>
<td>1.33</td>
<td>2.47</td>
<td>2.11</td>
</tr>
<tr>
<td>Proximity based</td>
<td>Nearest Neighbors</td>
<td>1.07</td>
<td>1.18</td>
<td>1.66</td>
<td>1.39</td>
</tr>
<tr>
<td>Probability based</td>
<td>Gaussian Process</td>
<td>3.66</td>
<td>2.48</td>
<td>4.10</td>
<td>3.76</td>
</tr>
<tr>
<td>Tree-structured</td>
<td>Decision Tree</td>
<td>0.66</td>
<td>0.95</td>
<td>1.36</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>Random Forest</td>
<td>0.88</td>
<td>0.90</td>
<td>1.42</td>
<td>1.03</td>
</tr>
<tr>
<td>Neural Network</td>
<td>1 layer</td>
<td>2.89</td>
<td>1.80</td>
<td>4.09</td>
<td>2.35</td>
</tr>
<tr>
<td></td>
<td>2 layers</td>
<td>1.90</td>
<td>1.77</td>
<td>3.78</td>
<td>2.36</td>
</tr>
<tr>
<td>Dimensionality Reduction</td>
<td>PCR</td>
<td>2.15</td>
<td>1.70</td>
<td>3.53</td>
<td>2.11</td>
</tr>
</tbody>
</table>
is to simultaneously determine the concentration levels of multiple macronutrients [5, 50, 59]. In SoilCares, we design a low-cost LED array to address the challenge. Specifically, we employ three spectral regions to detect N, P, and K concentrations separately. We utilize seven LEDs at 850, 950, 1150, 1200, 1300, 1550, and 1650 nm for N monitoring. We employ four LEDs at 400, 460, 470, and 525 nm for K monitoring. We utilize four LEDs at 620, 640, 660, and 720 nm for P monitoring. The selection of these spectral regions is strategically made to optimize the light attenuation by the target nutrient while limiting the impact of light attenuation by other nutrients [5, 28, 30, 49]. Fig. 4 illustrates the absorption properties of the target macronutrients N, P, and K at the selected wavelengths. We observe that the absorption of the targeted element is at least three times the absorption of other macronutrients and significantly more distinctive at the chosen wavelength, indicating the effectiveness of the chosen spectrum on soil macronutrient sensing.

3.4 Negligible-cost LoRa Transmission

The hardware of SoilCares consists of two main subsystems: RF-VNIR sensing node in soil and remote controller module. We built the controller module with a low-cost RF receiver RTL-SDR ($16) and a Raspberry Pi Zero ($5). To trigger the operation of the sensing node in the soil, an extra transmitter is required, which would increase the hardware cost. Instead, we propose to exploit existing hardware to enable negligible-cost LoRa signal transmission capability with delicate software programming. The key idea is to leverage existing RF-band signals in the current hardware for transmission. Fortunately, the Raspberry Pi has a programmable clock generator to drive its CPU up to 1 GHz, and the generated clock signal can be output to general-purpose input/output (GPIO) ports [2].

Fine-grained frequency control. Although sweeping the frequency to create chirp signals is straightforward, precise frequency control becomes the major challenge due to hardware limitations. The clock generator in Raspberry Pi is based on a 4GHz clock [54] and generates the target frequency through dividing the 4-GHz clock by a positive real number. The number is composed of a 12-bit integer part and a 12-bit fraction part [2]. Therefore, we can not obtain continuous target frequency due to the limited resolution. As shown in Fig. 7, for a 915 MHz channel with 125 kHz bandwidth, we can get only three discrete frequency points in the range by tuning the least significant bits, which is far from enough to generate a chirp signal.

To enable fine-grained frequency sweeping, we leverage a basic concept of signal processing: the instantaneous frequency is the time-domain derivative of the instantaneous phase:

\[ f(t) = \frac{1}{2\pi} \frac{d\phi(t)}{dt}. \] (10)

We further derive that the instantaneous frequency can be tuned by adjusting the “speed” of phase rotation, i.e., the accumulated phase rotation in a unit time. For example, we may quickly switch between the two discrete frequency points, 915 MHz and 915.052 MHz, half by half in a unit time. In this case, the accumulated phase rotation would equal that from a single-tone signal at the middle
while keeping the ratio of the two frequencies not changed. As a result, the spurs due to periodic switching, we randomize the switching order while keeping the ratio of the two frequencies not changed. As a result, the spurs are eliminated, and we obtain a clean single-tone wave.  

**Frequency harmonics reduction at GPIO.** Fine-grained frequency control is achieved with randomized frequency switching. However, the GPIO port of the Raspberry Pi is digital, i.e., it can only output “0” or “1.” In consequence, the output signal from the GPIO port is a square wave, which is the combination of the target frequency, 915.026 MHz. By adjusting the ratio of the two discrete frequencies, we obtain the target “fractional” frequency in the LoRa frequency band (902-928 MHz), as shown in Fig. 8. To eliminate the spurs due to periodic switching, we randomize the switching order while keeping the ratio of the two frequencies not changed. As a result, the spurs are eliminated, and we obtain a clean single-tone wave.

**Frequency harmonics reduction at GPIO.** Fine-grained frequency control is achieved with randomized frequency switching. However, the GPIO port of the Raspberry Pi is digital, i.e., it can only output “0” or “1.” In consequence, the output signal from the GPIO port is a square wave, which is the combination of the target frequency, 915.026 MHz. By adjusting the ratio of the two discrete frequencies, we obtain the target “fractional” frequency in the LoRa frequency band (902-928 MHz), as shown in Fig. 8. To eliminate the spurs due to periodic switching, we randomize the switching order while keeping the ratio of the two frequencies not changed. As a result, the spurs are eliminated, and we obtain a clean single-tone wave.

**Frequency harmonics reduction at GPIO.** Fine-grained frequency control is achieved with randomized frequency switching. However, the GPIO port of the Raspberry Pi is digital, i.e., it can only output “0” or “1.” In consequence, the output signal from the GPIO port is a square wave, which is the combination of the target frequency, 915.026 MHz. By adjusting the ratio of the two discrete frequencies, we obtain the target “fractional” frequency in the LoRa frequency band (902-928 MHz), as shown in Fig. 8. To eliminate the spurs due to periodic switching, we randomize the switching order while keeping the ratio of the two frequencies not changed. As a result, the spurs are eliminated, and we obtain a clean single-tone wave.
5 EVALUATION

In this section, we initially introduce the three evaluation metrics that will be utilized. This is followed by a detailed presentation of both in-lab and on-site experiments, which focus on monitoring soil moisture, nitrogen (N), phosphorus (P), and potassium (K) under a diverse range of experimental settings.

5.1 Evaluation Metric

For the soil moisture experiments in Sec. 5.3.1, we use the Mean Absolute Error (MAE) to facilitate the comparison with the state-of-the-arts [13, 22]. For the experiments on soil macronutrients, we leverage the coefficient of determination ($R^2$) as the primary criterion to enable comparison with existing studies [49, 50, 59]. The formula of $R^2$ is defined as Eq. 11:

$$R^2 = 1 - \frac{\sum_{k=1}^{m}(\hat{y}_k - f(x_k))^2}{\sum_{k=1}^{m}(y_k - \bar{y})^2}$$

where $\hat{y}_k$ is the ground truth of the $k$th macronutrient concentration and $m$ is the total number of experiment samples. The function $f$ is the adapted model function that transforms the input raw data $x_k$ to the predicted value of macronutrient concentration $f(x_k)$. $x_k$ is derived from the intensity of reflected light received at the photodiode, and $\bar{y}$ is the mean value of the macronutrient concentration ground truth. The coefficient of determination $R^2 = 1$ if the predicted value $f(x_k)$ exactly matches the observed value $y_k$. A negative value of $R^2$ indicates that the adapted model is worse than simply calculating the observed values’ mean.

We also leverage the mean square error of cross-validation (RMSECV) and root mean square error of prediction (RMSEP) as the evaluation metrics. These two criteria are considered as indicators of error in model predictions. The formulas of RMSECV and RMSEP are shown in Eq. 12:

$$RMSECV = \sqrt{\frac{1}{L_e} \sum_{l=1}^{L_e} (\hat{y}_l - y_{el})^2}$$

$$RMSEP = \sqrt{\frac{1}{L_p} \sum_{l=1}^{L_p} (\hat{y}_l - y_{pl})^2}$$

where $l$ is the sample number, $L_e$ and $L_p$ are the total number of samples in the validation and prediction groups, $y_{el}$ and $y_{pl}$ are the predicted concentration values of the target macronutrient from the validation group and prediction group.

5.2 Experiment Setting

Soil samples for in-lab evaluation. To validate the versatility of SoilCares across diverse soil types, we use three types of soil with distinct mixtures, shown in Fig. 10(h).

5.3 In-lab Evaluation

To evaluate the sensing performance across different moisture levels and soil types, we use transparent acrylic boxes (10 cm × 10 cm × 10 cm) to hold the above three types of soil samples with different moisture levels, spanning from the natural moisture level to saturation level (50%). For uniform soil moisture distribution, we cover the soil boxes with plastic wraps with uniform holes and let the specific amount of water drip through the holes uniformly. The high-end commodity soil sensor is used for comparison, shown in Fig. 10(e).
For macronutrient evaluation, in alignment with the procedures outlined by [43, 50], we use raw samples with a weight of 10 g, mixed with the target macronutrient and distilled water (an additional weight of 4 g), which are then transferred to a circle transparent acrylic board (38.5 mm in diameter) and covered by a cylindrical acrylic box (20 mm in depth) for experimentation.

We make 471 in-lab soil samples, encompassing all soil types and various concentrations. For the in-lab Soil NPK monitoring, 360 soil samples are prepared with macronutrients ranging from 0.1% to 1%. We also gather 81 samples, which are mixed with three concentration levels for each macronutrient: 0.2% for Low, 0.5% for Medium, and 0.8% for High. We establish three distinct levels for soil moisture: 10% for low moisture level, 20% for medium moisture level, and 28% for high moisture level to perform the concurrent soil N/P/K and moisture monitoring.

Soil and fertilization preparation for on-site study. To test SoilCares in practical scenarios, we conduct 108 on-site experiments at local farm fields. We select three field types for our study, as depicted in Fig. 11. To achieve the goal of assessing the effects of excessive soil fertilization, we adjust the macronutrient levels in the test fields to approximately 1.5 g/kg for nitrogen (N) [1, 52, 58], 1.5 g/kg for phosphorus (P) [6, 14, 58], and 2 g/kg for potassium (K) [52], while maintaining soil moisture levels between 20% and 45% throughout the duration of the experiment. The ground truth is measured after fertilization and full water infiltration [63] using the device in Fig. 10(e) and Fig. 10(f).

5.3.1 Performance of soil moisture and NPK monitoring.
We first evaluate SoilCares when only one soil substance (soil moisture, N, P, or K) is present in soil samples.

Soil moisture estimation. We conduct experiments to demonstrate the efficiency of our soil moisture monitoring module across a wide range of soil moisture levels (5–50% [13]) in agriculture. We repeat measurements 30 times for each soil sample and compute the mean output. We compare SoilCares with three baselines: the state-of-the-art LoRa-based method [13], a NIR-based method [42], and a high-end commodity device [3]. Fig. 12 shows the overall MAE across different moisture levels are 2.35%, 2.76%, 1.03%, and 2.54% for LoRa, NIR, SoilCares (LoRa + NIR), and the commodity device, respectively. SoilCares consistently achieves the highest accuracy for each moisture level and outperforms the high-end commodity sensor.

Soil NPK estimation. We then assess the ability of SoilCares to predict soil N, P, and K concentrations across three soil types. Table 2 summarizes the performance of SoilCares and three baselines [49, 50, 59]. Overall, SoilCares achieves a coefficient of determination of 0.811 for N, 0.803 for P, and 0.837 for K in soil sensing, comparable to existing methods that require soil pre-processing to enhance the prediction performance or expensive spectrometers. SoilCares only employs a COTS LoRa device, a few low-cost LEDs, and photodiodes to perform the function of sensing and data communication, which are more affordable and portable than the spectrometer-based method. Contrary to the methods utilized in [49, 50], our system avoids soil pre-processing (e.g., drying and grinding) and screening of collected soil samples. Finally, we observe that the existing LED-based method [49] achieves higher $R^2$ than SoilCares. However, the soil macronutrient concentration in their study (10–50%) substantially exceeds the concentration range specified in soil fertilization guideline [57], and there was no evidence that the system was evaluated under the practical macronutrient concentration level.

Impact of membrane across different soil types. For soil moisture estimation, we observe an MAE of 1.03%, reflecting a notable error reduction of 51.8% compared to the baseline outcome of 2.14% without membrane. This improvement is mainly attributed to the uniform surface provided by the membrane, facilitating consistent NIR reflection measurements across different soil types. For macronutrient sensing, Table 3 compares soil NPK estimation accuracy with and without the membrane. SoilCares achieves an $R^2 = 0.811$ for N, $R^2 = 0.802$ for P, and $R^2 = 0.837$ for K, significantly outperforming the performance without the membrane. To assess the performance of SoilCares across diverse soil types, we
Table 2: Experimental settings: we evaluate SoilCares performance and comparing with three related works

<table>
<thead>
<tr>
<th>Work</th>
<th>Device</th>
<th># of samples</th>
<th>Preprocessing</th>
<th>Concentration for N</th>
<th>Concentration for P</th>
<th>Concentration for K</th>
<th>Prediction Method</th>
<th>$R^2$ for N</th>
<th>$R^2$ for P</th>
<th>$R^2$ for K</th>
</tr>
</thead>
<tbody>
<tr>
<td>[50] Spectrometer</td>
<td>300</td>
<td>Required</td>
<td>0-3%</td>
<td>0-5%</td>
<td>0-0.3%</td>
<td>PCR / LS-SVR</td>
<td>0.87</td>
<td>0.99</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>[59] Spectrometer</td>
<td>280</td>
<td>Required</td>
<td>0.0032% - 0.0208%</td>
<td>0.0025% - 0.0343%</td>
<td>0.0041% - 0.0345%</td>
<td>PCR / LS-SVR</td>
<td>0.81</td>
<td>0.78</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>[49] LEDS</td>
<td>150</td>
<td>Required</td>
<td>10% - 50%</td>
<td>10% - 50%</td>
<td>10% - 50%</td>
<td>PCR / PLSR / LS-SVR</td>
<td>0.91</td>
<td>0.99</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>SoilCares</td>
<td>LEDS</td>
<td>Not Required</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Impact of membrane and diverse soil types on soil macronutrient monitoring

<table>
<thead>
<tr>
<th>Components</th>
<th>Calibration Method</th>
<th>$R^2$</th>
<th>RMSECV</th>
<th>RMSEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>N w/o membrane</td>
<td>LS-SVR</td>
<td>0.811</td>
<td>0.132</td>
<td>0.135</td>
</tr>
<tr>
<td>N w/o membrane</td>
<td>PCR</td>
<td>0.439</td>
<td>0.210</td>
<td>0.266</td>
</tr>
<tr>
<td>P w/o membrane</td>
<td>PCR</td>
<td>0.802</td>
<td>0.128</td>
<td>0.130</td>
</tr>
<tr>
<td>P w/o membrane</td>
<td>PCR</td>
<td>0.426</td>
<td>0.201</td>
<td>0.239</td>
</tr>
<tr>
<td>K w/o membrane</td>
<td>LS-SVR</td>
<td>0.837</td>
<td>0.095</td>
<td>0.103</td>
</tr>
<tr>
<td>K w/o membrane</td>
<td>PCR</td>
<td>0.436</td>
<td>0.216</td>
<td>0.278</td>
</tr>
</tbody>
</table>

Figure 13: Multi-modal resistance to RF/NIR bias

5.3.2 Performance of concurrent soil NPK and moisture monitoring. In this experiment, we assess the performance of our model, referred to as SoilCares, in scenarios where multiple confounding factors—including soil moisture, nitrogen (N), phosphorus (P), and potassium (K)—are concurrently present in soil samples. We partitioned the results into three groups based on their respective moisture levels: Low, Medium, and High, as illustrated in Sec. 5.2. Fig. 15 presents the SoilCares’s performance in relation to these four factors. Under medium soil moisture level, a sum absolute error of 2.36% for N, P, and K is achieved, which outperforms the sum error achieved under low moisture level (3.30%) and high moisture level (2.99%). This superior accuracy is mainly attributed to the lower error in moisture predictions around the medium level, which, after offsetting the effect of soil moisture, results in more accurate predictions for macronutrients.

The RMSE values under various moisture levels differ for each macronutrient element. We observed an RMSE of 0.139 for N, 0.150 for P, and 0.119 for K under the low moisture level. In the case of medium moisture level, the RMSE values are 0.101 for N, 0.109 for P, and 0.079 for K. Under the high moisture level, the RMSE values are 0.132 for N, 0.129 for P, and 0.137 for K. These results show that the RMSE of K varies significantly across the three settings. Apart from the factors related to moisture, we also observed that in the selected spectrum region for K monitoring, the chosen LEDs exhibit high sensitivity on N. This implies that the concentration level of N can influence the prediction of the concentration level of K to some extent [75]. In contrast, the spectrum regions selected...
for N and P do not exhibit sensitivity to other elements. Overall, our system achieves an average RMSE of 0.107 across three distinct soil moisture levels, each involving three macronutrients, indicating that SoilCares performs well in concurrently monitoring soil moisture and NPK levels.

5.4 On-site Study

Impact of device depth in the soil. We conducted on-site experiments at three depths: 0–10 cm, 10–20 cm, and 20–40 cm, as shown in Fig. 18(b). Most existing works [13, 20, 22] evaluated the performance of their systems at a depth up to 30 cm. The maximum testing depth (i.e., 40 cm) in this paper is determined based on the biological properties of plants. The roots of most plants are beneath the ground for about 20-30 cm, such as maize and celery [23, 45, 66]. We extend the sensing depth to 40 cm to cover more agricultural applications. Fig. 16(a) shows the RMSE of soil NPK at the three depths. We observe that the shallowest depth, 0–10 cm, yields the most accurate estimates, with an average RMSE of 0.112. As the depth increases, our system continues to deliver reliable results, with the RMSE always smaller than 0.182. Specifically, the predictions for N do not vary much across various depths, exhibiting an average RMSE of 0.145. For P and K, the average RMSE values are 0.147 and 0.129, respectively. These findings affirm that SoilCares can work effectively at various depths to meet the objectives of diverse applications in real-world scenarios.

Impact of terrains. We conducted on-site experiments to assess the effectiveness of SoilCares on various terrains, as depicted in Fig. 11. Our test sites included farmland with newly planted wheat, newly turned soil land, and land with wild grass. Fig. 16(b) shows that SoilCares is comparably effective across all three farmlands. Across the three terrains and macronutrients, the overall performance stands at a mean RMSE of 0.139. This result demonstrates that SoilCares is adaptable and functional across various soil types and top coverings.

Impact of weather conditions. To verify the consistency of our system’s performance under varying weather conditions, particularly under extreme weather (e.g., snow), we conduct experiments for a total of 60 hours. We start the experiment 12 hours after soil fertilization and water infiltration. After the inorganic fertilizers are applied, the macronutrient levels change rapidly in the next 72 hours and need to be closely monitored to avoid negative impacts on plants and the potential risk of contaminating groundwater. After 72 hours, the macronutrient levels change much more slowly. Fig. 16(c) illustrates the changes in absolute error and the temperature fluctuations. Notably, there are two significant increases in error during the 36-48th hours and 72nd hour intervals, influenced by abrupt drops in soil temperature. In the 36-48th hours, the soil temperature dropped below 0 °C. There was snowfall in the 72nd hour. These weather changes caused the moisture in the soil and the
membrane to freeze, resulting in an RMSE exceeding 0.20. Despite these challenges, our system still produces reliable predictions with an RMSE lower than 0.165 for all other cases when the soil temperature is higher than 0 °C, the weather condition for SoilCares to provide reliable monitoring results.

**Impact of soil moisture estimation on soil NPK estimation.** Since the accuracy of soil moisture prediction is closely linked to the accuracy of macronutrient estimation, we investigate the impact of soil moisture estimation errors on soil macronutrient prediction. In this experiment, we focus on soil N prediction. We manually add noise to our soil moisture module to increase the errors in moisture sensing. Fig. 16(d) shows the RMSE of soil N prediction under various moisture estimation errors. The prediction error significantly escalates as moisture estimation error increases. Reducing moisture estimation error from 5% to 1% can significantly increase the prediction accuracy of nitrogen. This indicates that accurate soil moisture estimation plays a critical role in the performance of macronutrient prediction. The achieved soil moisture estimation accuracy in our system is adequate for reliably monitoring soil macronutrients [49, 50, 59]

**Performance of negligible-cost LoRa transmission.** We also evaluate the performance of the proposed negligible-cost LoRa transmission of the remote controller in the on-site study. The sensing node is buried in the soil, with the membrane and RF antennas positioned outside the box to ensure direct contact with the soil, while the circuits are placed inside the box. The sensing node is buried 60 cm below the soil surface at different locations under different soil moisture levels (14%, 23%, and 55% by adding water). Then, we move the controller away from the sensing node at a step size of 10 m from 0 m (right above the sensing node) to 80 m, as shown in Fig. 18(a). To evaluate the performance of the controller for LoRa signal transmission, we adopt the default parameter setting for LoRa transmission, i.e., a central frequency of 915 MHz with a channel bandwidth of 125 kHz, a spreading factor of 12, and a 4/5 coding rate. The controller is programmed to send 100 LoRa packets. At the sensing node side, the commodity LoRa node acts as a receiver and decodes the received LoRa packets. Since the software [9] used in the sensing node only outputs when a packet is correctly decoded, we cannot obtain the bit error rate. Therefore, we plot the reported signal-to-noise-ratio (SNR) of those packets correctly decoded, as shown in Fig. 17. When the remote controller is right above the under-soil sensing node, the mean SNR across different moisture levels is all above 12 dB. Even in the most challenging case under a moisture level of 55%, we can still achieve successful packet reception at a distance of 80 m, where the mean SNR drops to -15 dB. Based on this result, the proposed negligible-cost LoRa transmission can achieve a transmission range larger than 80 m in real-world farmland, covering an area over 20,000 m².

### 6 RELATED WORK

#### 6.1 Light-based soil macronutrient sensing

Reflectance spectroscopy has found extensive usage in soil analysis for a large range of elements, such as nitrogen [5, 28], phosphorus [18, 49], potassium [16, 30, 43], and other critical elements like organic carbon [31]. The majority of accurate soil element measurements are conducted under laboratory settings using expensive and bulky spectrometers, which provide high-quality spectral resolution for analyses [29, 50, 59]. These lab-oriented studies often involve pre-processing steps such as drying and grinding the soil to mitigate the effects of confounding factors like soil moisture and particle size. This helps ensure higher measurement accuracy and
applicability to different soil types or textures [50, 59]. However, the high cost of spectrometers and the need of pre-processing severely limit the wide adoption of reflectance spectroscopy for everyday use. Recent studies have tried to circumvent these issues by adopting a combination of LEDs and photodiodes as an alternative to spectrometers [5, 49, 75]. Nevertheless, these works have not fully addressed the problem of generalizing across different soil types and still require tedious pre-processing. Moreover, they do not offer simultaneous prediction of multiple elements.

6.2 RF-based soil moisture sensing
RF-based soil moisture sensing is an emerging field. To replace the expensive ultra-wide-band ground penetrating radar (GPR) [4, 37], RF-based soil moisture sensing has been proposed to achieve low-cost sensing. A variety of RF signals have been exploited for soil moisture sensing including WiFi [20], RFID [70], LoRa [13] and LTE [22]. The WiFi-based solution proposes to use WiFi channel state information (CSI) to realize both moisture and salinity sensing. The RFID-based approach leverages RFID signal attenuation for moisture sensing. It is limited to container cases since the RFID tag needs to be attached to the outer surface of the container. In comparison, our proposed system can be applied to much broader application scenarios and achieves a higher accuracy (1% error vs. 3% error). The LoRa- and LTE-based solutions mainly focus on extending the range of soil moisture sensing and they are not capable of sensing macronutrients. A recent work [36] achieves a low moisture estimation error (1.1%) leveraging customized RF signals with a large GHz-bandwidth and expensive high-end software-defined radio platform ($8000). In comparison, our system achieves a similar accuracy of 1% with low-cost hardware.

7 DISCUSSION

On-site soil macronutrients and moisture sensing. The primary distinction between laboratory and field experiments lies in the additional variables present in the wild that are not encountered in the lab, such as soil temperature and surface vegetation. It is essential to maintain the temperature above 0 degrees Celsius to ensure reliable sensing. Lower soil temperatures can alter the state of water, complicating the measurement of macronutrients and degrading the sensing performance. Moreover, stones and sharp objects could damage the membrane’s surface, affecting the VNIR reflection. Furthermore, improper setup of the system, such as having large air gaps above the membrane, can also negatively impact system performance.

Over 50% soil moisture. Soil moisture levels in agriculture usually fall in the range 5% to 45% and the proposed system works well in this range. We also notice that in extreme cases (e.g., rice paddy), the soil moisture can be higher than 50%. In this case, the principles of the proposed system still hold. However, the high moisture can attenuate RF signals significantly, reducing the sensing range.

8 CONCLUSION

In this paper, we present SoilCares, a system capable of simultaneously measuring soil moisture and soil macronutrients using RF-VNIR sensing. To accurately sense soil moisture and macronutrients across diverse soil types and textures, we design a novel membrane to provide a uniform reflectance spectroscopy for Vis-NIR sensing. Meanwhile, by leveraging the COTS LoRa hardware and low-cost LEDs and photodiodes, we propose a multi-modal model combining reflectance spectroscopy and RF sensing for soil moisture and macronutrient sensing. Extensive experiments show that SoilCares can achieve a root-mean-square error of 0.138 on soil macronutrient monitoring and 1% mean absolute error on soil moisture monitoring in complex real-world scenarios.

REFERENCES

[19] Digigkey. 2023. 915MHz Bandpass Filter. https://www.digigkey.com/en/products/filters/rf-filters/84477-N4lgICboEwdlDGBmBDaDgANZwKryQ8KABjRYs82ATq5YAO EAXQ1AAXK34CCal3gDaA5iA%22BMORpRyQ050m5e4pCAYMN 5CFR0usKxGlHi50a


