Geotech Geol Eng (2025) 43:318
https://doi.org/10.1007/s10706-025-03293-x

ORIGINAL PAPER

Soil Moisture Prediction in Pavement Layers Using LSTM Neural Networks

Amir Tophel · Thi Mai Nguyen · Jeffrey P. Walker · Jayantha Kodikara

Received: 18 May 2025 / Accepted: 7 July 2025

Abstract Moisture content in geomaterials critically impacts road construction. Optimising the Optimum Moisture Content (OMC) during compaction reduces energy usage and construction costs, but materials often require adjustments upon delivery to reach OMC. The dry back phase is essential for efficiently allowing the pavement to release trapped moisture, preventing surface issues such as moisture resurfacing, aggregate punch-in, and premature stiffness loss in pavements. This study investigated temporal soil moisture (SM) variation within the compaction layer under varying environmental conditions such as physical soil temperature (T_{soil}), net radiation (Rn), initial SM, and bulk density (ρ_b). A Long Short-Term Memory (LSTM) model was developed and compared with a traditional three-layer water balance model to predict SM variations. The LSTM model showed superior accuracy, with a maximum Root Mean Square Error of 0.20% for Volumetric Moisture Content and 0.12% for Gravimetric Moisture Content. Integrating environmental data and site-specific soil characteristics with a data-driven model significantly advances moisture management in road construction, enhancing compaction and pavement durability.

Keywords Soil moisture · Pavement layers · Machine learning · LSTM model · Optimum moisture content · Dry back

Abbreviations

S	Degree of Saturation (%)
ρ_b	Bulk density (kg/m^3)
T_{soil}	Soil temperature (°C)
Rn	Net radiation (W/m^2)
VMC	Volumetric Moisture Content (v/v) (%)
GMC	Gravimetric moisture content (w/w) (%)
NDG	Nuclear Density Gauge
RMSE	Root Mean Squared Error
LSTM	Long short-term memory
SM	Soil moisture (%)
OMC	Optimum Moisture Content (w/w) (%)
MDD	Maximum Dry Density (kg/m^3)

1 Introduction

Compaction in road construction involves the mechanical compression of geomaterials using large rollers to achieve the designated density, enhancing the material's strength and stiffness. This process is essential for preventing settlement or rutting

A. Tophel (✉) · T. M. Nguyen · J. P. Walker · J. Kodikara
Department of Civil Engineering, Monash University, Clayton Campus, VIC 3800, Australia
e-mail: amirtophel@gmail.com; amir.tophel@monash.edu

T. M. Nguyen
e-mail: thi.nguyen9@monash.edu

J. P. Walker
e-mail: jeff.walker@monash.edu

J. Kodikara
e-mail: jayantha.kodikara@monash.edu

Published online: 23 July 2025

post-construction (Briaud 1997). Geomaterials are typically compacted in layers, with each layer's thickness ranging from 15 to 20 cm. Achieving Maximum Dry Density (MDD) at the Optimum Moisture Content (OMC) is vital for efficient soil compaction, as it reduces construction costs by minimising the required compaction energy (Kodikara et al. 2018).

However, the materials delivered to construction sites often do not meet the OMC, either too wet or too dry, necessitating adjustments such as drying or adding water. Sometimes, even when the materials possess adequate moisture content during compaction, a time lag in spreading the material and placing it can lead to drying out, preventing proper compaction (Chen et al. 2013). Therefore, knowing the time lag and initial moisture content is essential to predict how much water is needed to reach OMC.

Furthermore, an integral part of the construction process is the dry back phase, where the geomaterials undergo drying to a target Degree of Saturation (S) after compaction. This step is crucial for increasing the hardness and stiffness of the pavement layers (Athmarajah et al. 2024). The effective dry back allows the pavement to release moisture efficiently, preventing issues like moisture surfacing below the seal, which can cause aggregates to punch into the pavement and reduce stiffness (Lim, Hore-Lacy et al. 2014). The accurate prediction of the time required to reach the target S during dry back is essential for optimising the placement of subsequent layers and minimising the risk of improper moisture levels.

The measurement of soil moisture (SM) of these geomaterials for both achieving OMC during compaction and monitoring dry back post-compaction in the field is also vital, and typically it is done using laboratory moisture determination or the Nuclear Density Gauge (NDG) technique (Salem 2004). However, these traditional measurement techniques can be limited in accuracy and coverage and are often time-consuming or hazardous. Nguyen et al. (Nguyen 2023; Nguyen et al. 2023; Nguyen et al. 2023) developed a method for real-time measurement of unbound pavement layer SM using an L-Band radiometer. This technique mainly measures near-surface SM, which is only a fraction of the compacted layer's 15–20 cm thickness. Understanding the initial SM condition allows for modelling the variation of SM within the entire depth of the layer over time, which is crucial for managing water requirements for optimal compaction and effective dry back.

The temporal variation of SM is influenced by numerous factors, including physical soil temperature (T_{soil}), net radiation (Rn), amount of water in the soil (known as initial SM condition), and bulk density (ρ_b) Tamboli 1961; English et al. 2005; Wang et al. 2007; Gao 2011). Net radiation (Rn) is a key environmental factor, defined as the difference between the incoming and outgoing radiation at the soil surface. It influences SM by affecting evaporation and soil temperature, thus playing a critical role in water movement through soil.

The following points outline some of the physical controls on water movement through soil:

- Water viscosity decreases with increasing T_{soil}, which allows more water to flow through the soil matrix, resulting in a faster decrease in SM (Broadbent 2015). In addition, higher T_{soil} results in a higher evaporation rate, further accelerating SM reduction (Rengasmy 2009).
- Rn is one of the important components of the Earth's surface energy balance and is a major driving force for the process of evaporation (E) of the soil surface (Wang et al. 2007). SM is directly related to the evaporation process and is usually controlled by the moisture in the upper layer of the soil profile (Verstraeten et al. 2008).
- The amount of water in the soil influences infiltration, runoff and evapotranspiration rate (Grayson et al. 1997).
- Soil compaction caused by heavy machinery or compactors increases ρ_b(Rafiq 1990; Hassan and Gregory 1999), which results in changes in infiltration, water holding capacity, water movement, and the ability for soil to shrink (Dec et al. 2008).

To predict SM at deeper levels based on near-surface observations, a variety of methods are employed, such as data assimilation (Montaldo et al. 2001, 2007), exponential filter (EF) (Ford et al. 2014a, b), or data-driven methods, such as machine learning (Kolassa et al. 2018). The advantage of traditional models is that they capture the basic physics of the moisture flow (Ali et al. 2015), but may not be able to capture the complex process of moisture flow or soil variability. Some calibration may be required, particularly for parameters

like soil porosity and hydraulic conductivity under site-specific conditions such as high clay content or extreme compaction (Reichle et al. 2008). Moreover, in many instances, such approaches are subject to strict statistical assumptions, and data requirements often employ linear and additive modelling approaches that may be inconsistent with natural processes (Clapcott et al. 2013).

Field and modelling studies have revealed strong nonlinear correlations between near-surface and subsurface soil moisture (commonly referred to in agricultural studies as root zone soil moisture, which refers to the moisture content within the depth accessible to plant roots, typically 0–100 cm), challenging conventional statistical and modelling methods (Ford et al. 2014a, b). While this terminology originates from agronomy, we reference it here to draw conceptual parallels with moisture behaviour in the compacted sublayers of pavements, where unsaturated flow processes are similarly governed by capillarity, evaporation, and soil structure. This analogy highlights the need for models that can accurately capture nonlinear moisture dynamics in compacted layers (typically 15–20 cm thick) used in road construction, where conventional approaches may fall short.

Advancements in machine learning techniques offer promising solutions to overcome the limitations of traditional models. For instance, machine learning-based random forest models perform better than physical models based on Richard's equation for predicting soil matric potential in the root zone (Gumiere et al. 2020). In other studies, artificial intelligence has introduced sophisticated models that significantly enhance the predictive accuracy of soil moisture (Ahmed et al. 2021; Filipović et al. 2022; Park et al. 2023a, b; Huang et al. 2024; Wang et al. 2024). For instance, studies have shown that machine learning algorithms such as Random Forest and Long short-term memory (LSTM) have been effectively utilised in various settings, including agriculture and subgrade materials for roads, railways, providing reliable predictions for moisture content at various depths and in real-time (Chen et al. 2021, Brakhasi et al. 2024, Geng et al. 2024, Dushmantha et al. 2025). Notably, physically-guided LSTM models, as demonstrated by Wang et al. (2024) can enhance predictive accuracy while preserving interpretability in soil moisture modeling, offering valuable insights for adapting these methods to road construction contexts.

Recently, Ahmed and Karimpour (2025) used Random Forest and Gaussian Mixture Models to successfully model subgrade moisture in cold regions, validating the potential of machine learning for capturing seasonal SM dynamics under varying thermal regimes.

Despite these advancements, the specific application of such AI techniques to road construction remains underexplored, particularly for predicting deep-layer soil moisture relevant to compaction. This research seeks to bridge this gap by employing advanced LSTM models tailored to capture the complex dynamics of soil moisture across different compaction layers, which significantly improves over traditional methods. Our approach leverages a novel combination of environmental data and site-specific soil characteristics to provide more accurate and timely predictions for efficient compaction. This study advances the methodological framework for soil moisture prediction in road construction by integrating more robust and real-time data-driven insights. It contributes to the broader field of construction engineering by improving process efficiencies and material optimisation. This comprehensive approach and novel application of LSTM models demonstrate a significant enhancement in modelling capabilities, addressing the critical need for precise moisture control in road construction projects and setting a new standard in the application of AI in civil engineering.

This paper uses the term "layer" to indicate different depths within a single compacted lift. Various factors, including T_{soil}, Rn, initial SM condition, and ρ_b. They were collected and analysed for their influence on the temporal variation of SM in different layers for two different materials under different environmental conditions. This dual-material approach was designed to enhance the generalizability of the findings, providing insights into the behaviour of different soil types, each representative of typical soil types encountered in pavement construction. Accordingly, the near-surface SM observations were utilised to predict the subsurface SM of deeper layers using a machine learning-based LSTM, considering all these factors as potential input variables. For comparison, a traditional water balance model was also employed to predict the temporal variation of SM, and the results were compared with those from the LSTM model.

2 Methodology

2.1 Experimental setup

Temporal variation of Volumetric Moisture Content (VMC) data at different compaction densities, depths, and initial SM conditions was recorded using Stevens Hydra Probe SM sensors (Fig. 1a). VMC refers to the volume of water per unit volume of soil (typically expressed as cm^3/cm^3 or %). The Stevens Hydra Probe sensors used in the setup provide real-time, direct measurements of VMC and soil temperature (T_{soil}) simultaneously (Stevens 2022).

Two experiments, each with two different dates, were performed in two environments: an indoor soil box (Experiment A) and an outdoor test-bed (Experiment B). For both the soil box (4 m × 7.5 m) and test-bed (5 m × 12 m), SM sensors were installed on one side, as shown in Fig. 1b. Two different materials, consisting of a 20 cm (indoor; soil sample 1) and 15 cm (outdoor; soil sample 2) layer thickness, were used, with the characteristics of these two samples shown in Fig. 2 and Table 1. Soil sample 1 has higher density (2.08 vs. 1.85 t/m^3) and lower OMC. Accordingly, 4 and 3 SM sensors were used to record the vertical profile of the VMC and T_{soil} at 5 cm depth intervals at the soil box and test-bed, respectively (Fig. 1c). These SM sensors

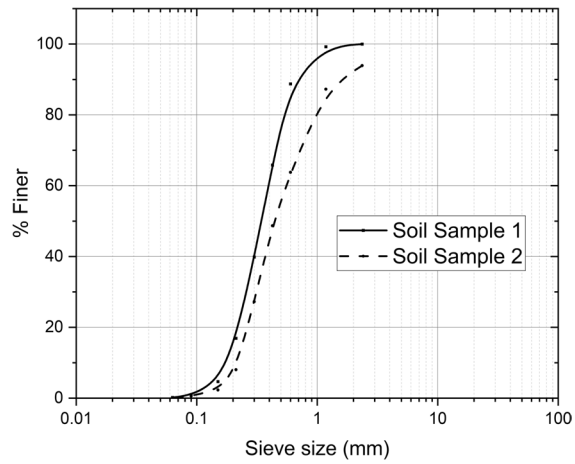

Fig. 2 The grain size distribution of the two soil samples used in this study

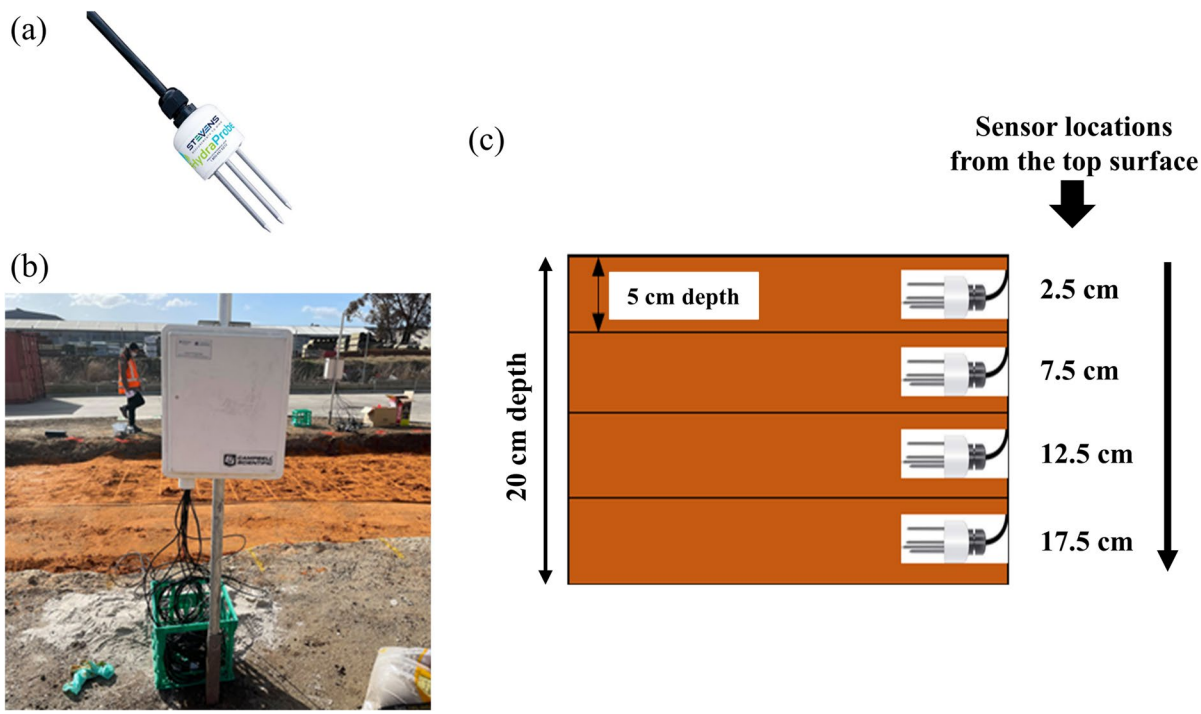

Fig. 1 Photographs showing the Stevens Hydra Probe SM sensor (**a**) and the station located at the test-bed (**b**) with four sensors inserted into the soil at four different depths as shown in the diagram (**c**)

Table 1 Geotechnical properties of the samples used in this study

Property	Soil Sample 1	Soil Sample 2
Specific Gravity (G_s)	2.61	2.70
Optimum Moisture Content (OMC), modified Proctor (%)	9.3	8.0
Maximum Dry Density (MDD), modified Proctor (t/m^3)	1.85	2.08
Mean Particle Diameter (D_{50}) (mm)	0.45	0.32
Percentage passing the No. 200 sieve (%)	0.2	0.1

continuously measured VMC and T_{soil} at 2-min sampling steps.

2.1.1 Soil sample Preparation

The materials were placed using a bobcat/mini-loader and then compacted using a double-drum roller (Fig. 3). After compaction, the density was measured using the NDG technique. The SM sensors were installed at the middle depth of each layer.

The indoor datasets from Experiment A (Day 1 and Day 2) were used to analyse the temporal variation of SM without exposure to rapid changes in ambient conditions. Experimental data included VMC and T_{soil} values at four different depths, being 2.5 cm (representing the top layer 0–5 cm), 7.5 cm (representing the second layer 5–10 cm), 12.5 cm (representing the third layer 10–15 cm), and 17.5 cm (representing the fourth layer 15–20 cm). The first dataset was collected from 3.04 pm on 02-Dec to 7.52 am on 03-Dec 2020, while the second dataset was recorded only during the daytime from 8.28 am to 6.32 pm on 03-Dec 2020.

The outdoor datasets from Experiment B (Day 1 and Day 2) were used to analyse the temporal variation of SM considering the influence of ambient conditions (i.e., *Rn*) and soil compaction (i.e., ρ_b). Each data set included VMC data at three different depths: 2.5 cm (representing the top layer 0–5 cm), 7.5 cm (representing the second layer 5–10 cm), and 12.5 cm (representing the third layer 10–15 cm). VMC was the preferred metric for this study because the Stevens Hydra Probe sensors used in the setup provide real-time, direct measurements of VMC, but for practical purposes, the VMC data were converted to Gravimetric Moisture Content (GMC) using the density (ρ_b) information from the NDG. GMC refers to the water mass per unit of dry soil mass expressed as g/g or %. The first outdoor dataset was collected on 29-Jul 2021, while the second was recorded on 02-Dec 2022. The *Rn* and evaporation (E) data were taken from the Wombat station (the nearest weather station to the experimental setup) at 30-min sampling steps, available from the OzFlux website (OzFlux 2022). Information on testing time and measurement conditions for each experiment is shown in detail in Table 2.

(a) **(b)**

Fig. 3 Photographs showing the roller used to compact the soil at (**a**) the soil box and (**b**) the test-bed

Table 2 Details of testing time and measurement conditions in each experiment. The tick indicates that the data was used on the respective date

Environment	Date	Experiments	Time (min)	T_{soil} (°C)	Rn (W/m^2)	ρ_b(g/cm^3)	Depth (cm)
Indoor (soil box) using soil sample 1	02-Dec, 2020	Day 1 of Experiment A	√	√			√
	03-Dec, 2020	Day 2 of Experiment A	√	√			√
Outdoor(test-bed)using soil sample 2	29-Jul, 2021	Day 1 of Experiment B	√		√	√	
	02-Feb, 2022	Day 2 of Experiment B	√		√	√	√

2.2 Model Development

2.2.1 Long Short-Term Memory Model

Long short-term memory (LSTM) is one of the machine learning tools used to capture long- and short-term dependencies and maintain information over extended sequences (Hochreiter and Schmidhuber 1997; Tophel et al. 2023a, b, c). This makes LSTM particularly suitable for modelling time-series data with complex temporal dependencies. Several studies have explored the use of LSTM for predicting moisture at deeper depths in a range of different fields (Fang et al. 2017; Han et al. 2021; Gao et al. 2022; Yinglan et al. 2022; Park et al. 2023a, b). These studies leverage the strengths of LSTM in handling sequential data and have shown promising results in accurately capturing SM dynamics. Furthermore, the LSTM model has demonstrated superior predictive performance to other machine learning models (i.e. artificial neural network) when predicting SM at different depths (Han et al. 2021).

The LSTM algorithm comprises four components: the forget gate, input gate, temporary cell state, and output (Fig. 4). The forget gate in LSTM determines the amount of information from the previous block that should be preserved. The input gate plays a role in determining which new information should be stored in the cell. Lastly, the output gate selects the final output value from the information stored in the cell. The LSTM algorithm operates on an input data sequence (X_t) to generate the outcome (O_t) by iterating through

$$f_t = \sigma(W_f(h_{t-1}, x_t) + b_f) \tag{1}$$

$$i_t = \sigma(W_i(h_{t-1}, x_t) + b_i) \tag{2}$$

$$o_t = \sigma(W_o(h_{t-1}, x_t) + b_o) \tag{3}$$

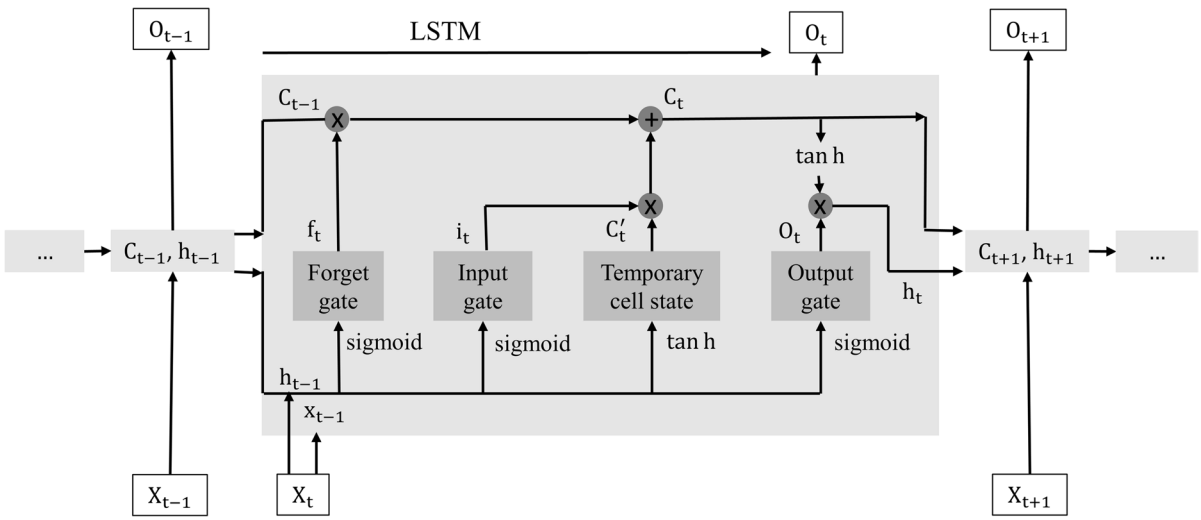

Fig. 4 Structure of an LSTM model

$$C_t\prime = \tanh(W_C(h_{t-1}, x_t) + b_C) \tag{4}$$

$$C_t = f_t \times C_{t-1} + i_t \times C_t\prime \tag{5}$$

$$h_t = O_t \times \tanh(C_t), \tag{6}$$

where σ represents a nonlinear activation function; W_f, W_i, W_o, and W_C denote the weight values associated with the forget gate, input gate, output gate, and memory cells, respectively; h_t represents the output data from the previous cell; x_t represents the current input data; b_f, b_i, and b_o are the bias vectors corresponding to each gate; C_t refers to the cell state in the LSTM network, which retains memory of the sequence and $C_t\prime$ represents the state of any cell generated from the activation function.

The dataset from Day 1 and Day 2 of Experiment A was divided into two parts for this study. The first 80% of the data from all four layers in Day 1 of Experiment A was used to train the model, while the remaining 20% was used for testing. After developing the model, the entire dataset from Day 2 of Experiment A was used as the validation set. For outdoor analysis in Day 1 and Day 2 of Experiment B, the same approach of using the first 80% of the data for training and the remaining 20% for testing was employed. The specific details of the data division are presented in the respective result sections. During the training process, the first 10 data points were utilised for the input gate, while the subsequent 2 data points were employed for the output gate. This pattern continued until all of the selected training data was utilised. The model hyperparameters (number of LSTM layers, number of neurons in each layer, learning rate) were found using the GridSearch algorithm available in the scikit-learn library (Pedregosa, Varoquaux et al. 2011, Tophel et al. 2022a, b, c, Tophel et al. 2023a, b, c). The grid search was conducted over the following ranges: number of hidden layers from 1 to 5, number of neurons per layer from 5 to 15, and learning rate from 0.1 to 0.5. The model performance for each hyperparameter combination was evaluated using fivefold cross-validation, scoring based on the negative mean squared error. This ensured consistent evaluation across folds and avoided overfitting to a specific data split.

The optimal configuration—2 hidden layers, 10 neurons per layer, and a learning rate of 0.1 was selected based on achieving the lowest average error across the folds. These values were then applied consistently in all experiments presented in this study, balancing model complexity and predictive performance.

The contribution of each input parameter to the model is shown by the variable importance (*VI*), otherwise known as the sensitivity analysis of input parameters. The sensitivity of each parameter (j) was assessed in this study by examining the change in the model output ($\Delta\widehat{Y}_j$) when that specific input parameter was changed while keeping the other input parameters constant at their mean values. Each input was altered from its maximum (j_{max}) to minimum (j_{min}) value from the database, and the *VI* calculated as the per cent of each absolute output difference for each parameter such that

$$\Delta\widehat{Y}_j = abs(f(j_{max}) - f(j_{min})) \tag{7}$$

$$VI_j = \frac{\Delta\widehat{Y}_j}{\sum_{j=1}^{j=n}\Delta\widehat{Y}_j}. \tag{8}$$

Pearson's correlation coefficient (Sedgwick 2012) was used to measure the linear correlation between the two variables (i.e., SM and T_{soil}) in this study.

2.2.2 Water Balance Model

A water balance model estimates the amount of water that enters and leaves a given system over a specified period. It is based on the principle of conservation of mass and energy, where the amount of water that enters a system (through precipitation or irrigation) is balanced against the amount of water that leaves (through runoff, evapotranspiration, or drainage) and change in storage (Zhang et al. 2002). In this study, a simple three-layer water balance model was designed, as shown in Fig. 5.

The first layer represents the top near-surface layer (i.e. 0–5 cm depth), which experiences evaporation (E) to the atmosphere (there was no precipitation or watering during the experiment conducted, and transpiration was zero as there were no plants) and redistribution to/from the second layer. Redistribution of SM refers to water movement within the soil from low matric suction (typically high SM) areas to high matric suction

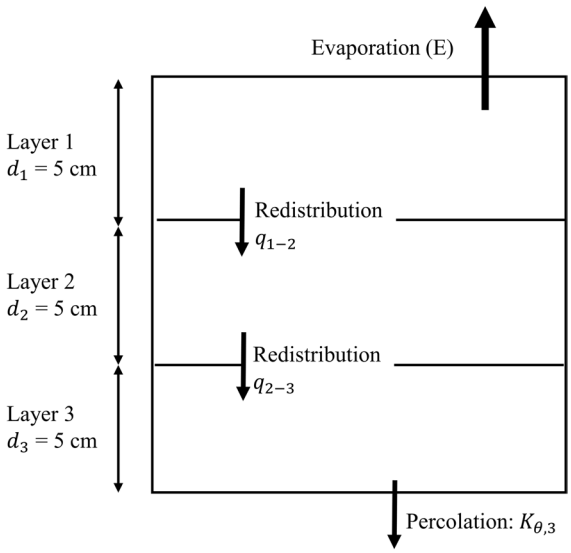

Fig. 5 Concept of the simple three-layer water balance model

(typically low SM). The water balance equation for this layer can be expressed from continuity as

$$\theta^{n+1} = \theta^n + \frac{\Delta t}{d_1}\left[-E^n - q_{1-2}^n\right], \tag{9}$$

where θ is the moisture content, n is the time step, d_1 is the layer depth, and Δt is the time step size. The flux from layer 1 to layer 2 (q_{1-2}) is

$$q_{1-2} = K_{\theta,1}\frac{\psi_{m,1} - \psi_{m,2}}{0.5(d_1+d_2)} + K_{\theta,1}, \tag{10}$$

where the flux out of the bottom of layer 1 is $K_{\theta,1}^n$, whereby K_θ can be calculated based on the Brooks and Corey equation (Brooks 1965) as

$$K_\theta = K_s\left(\frac{\theta - \theta_r}{\phi - \theta_r}\right)^c \tag{11}$$

$$c = 2b + 3, \tag{12}$$

where K_s is the saturated hydraulic conductivity, θ_r is the residual soil water content, and ϕ is the soil porosity. The variables $\psi_{m,1}$ and $\psi_{m,2}$ are the matric potential of layer 1 and layer 2, respectively, and can be calculated from

$$\psi_m = \psi_a\left(\frac{\theta - \theta_r}{\phi - \theta_r}\right)^{-\phi}, \tag{13}$$

where ψ_a is air entry pressure.

The second layer represents the subsoil layer (i.e. 5–10 cm depth) below the topsoil. The water balance equation for this layer can be expressed as

$$\theta^{n+1} = \theta^n + \frac{\Delta t}{d_2}\left[{q_{1-2}}^n - {q_{2-3}}^n\right], \tag{14}$$

where the flux from layer 2 to layer 3 (q_{2-3}) is

$$q_{2-3} = K_{\theta,2}\frac{\psi_{m,2} - \psi_{m,3}}{0.5(d_2+d_3)} + K_{\theta,2}, \tag{15}$$

where the flux out the bottom of layer 2 is $K_{\theta,2}^n$ and $\psi_{m,3}$ is the matric potential of layer 3.

For the third layer (i.e. 10–15 cm depth), the water balance equation can be expressed as

$$\theta^{n+1} = \theta^n + \frac{\Delta t}{d_3}\left[{q_{2-3}}^n - K_{\theta,3}^n\right], \tag{16}$$

where the flux out the bottom of layer 3 is $K_{\theta,3}^n$ under the assumption of gravity drainage. However, in this study, the flux out of the bottom of layer 3 ($K_{\theta,3}^n$) was set equal to 0 because the natural soil beneath the third layer displayed very low hydraulic conductivity.

In this study, the dataset from Day 2 of Experiment B was used exclusively to analyse the water balance model, as it was the only dataset containing density information for all three layers. The equations from the developed water balance model can be solved explicitly for each time step (an analysis time of 2 min was used to match with the time step of the experimental data and LSTM model), giving a new moisture content for each layer of the profile. Accordingly, it is possible to model the water balance in this three-layer soil system and predict the change in SM over time.

3 Results and Discussion

3.1 The Temporal Variation of Soil Moisture at Different Layers for the Indoor Environment

The data from Day 1 and Day 2 of Experiment A were used to study the temporal variation of SM without the effect of ambient conditions, as they were indoor experiments. Moreover, these two datasets provided an opportunity to examine the effect of SM variation separately during night and daytime.

Figure 6 (a) and (b) show the change of T_{soil} for the four different layers (0–5 cm, 5–10 cm, 10–15 cm, and 15–20 cm) for Day 1 and Day 2 of Experiment A. Regardless of whether the data were observed at night or day, T_{soil} near the surface (i.e. 0–5 cm) always showed the most rapid change, whereas deeper layers had a smaller and more gradual change.

According to Fig. 6, the SM variation of the corresponding four layers showed different trends over time. In general, the change of SM for the near-surface layer (0–5 cm) decreased with time due to redistribution to the deeper layer and evaporation from the soil-air interface. The second (5–10 cm) and third (10–15 cm) layers showed the same trend, with SM being mostly constant, which can be explained by approaching the equilibrium profile. In contrast, the SM at the fourth layer (15–20 cm) increased steadily over time because of percolation from the soil above, with an impermeable layer below.

Since the soil box was constructed on top of a waterproof membrane, water accumulated at the bottommost layer in the soil box from gravity drainage of the soil layers above. It was additionally observed that the SM variation of the different layers depended on the initial SM values. When the SM was low at the start, the change in SM over time was greater than for the same soil with a higher initial SM. When the initial moisture content was low, the soil had a higher water-holding capacity and could quickly absorb and retain water. Therefore, any additional water input is more likely to cause a greater change in SM over time.

Conversely, when the initial moisture content was high, the soil already had a lower water holding

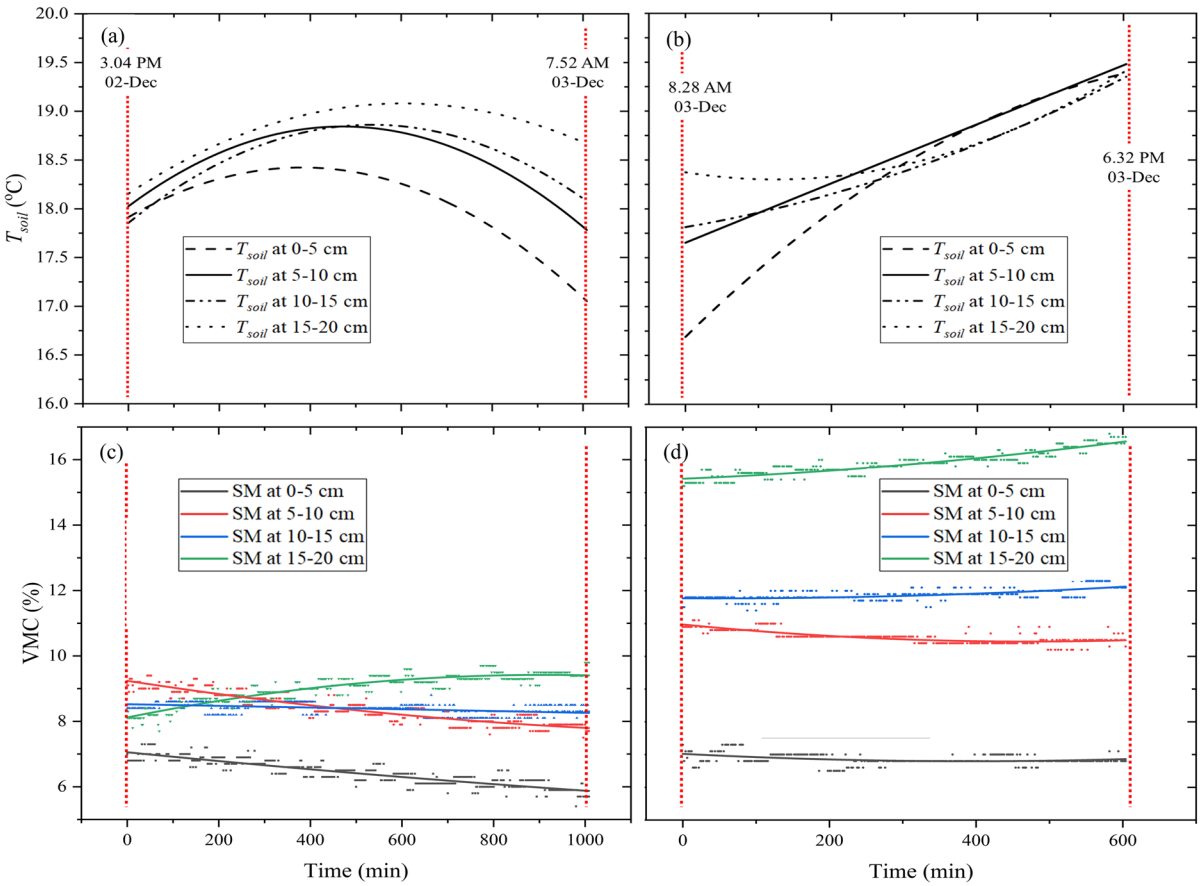

Fig. 6 T_{soil} data for the four different layers (0–5 cm, 5–10 cm, 10–15 cm, and 15–20 cm) as recorded in **a** Day 1 of Experiment A and **b** Day 2 of Experiment A. The lower parts are as for T_{soil} but for SM in **c** Day 1 of Experiment A and **d** Day 2 of Experiment A. The red dotted lines indicate the start and end times of recorded data

capacity available. It was more likely to reach its maximum, so additional water input may not cause a significant change in SM. In other words, the soil with a higher initial moisture content is already closer to saturation, limiting the potential for further water absorption. This is why soil with low initial moisture content can experience greater changes in moisture content over time than soil with higher initial moisture content. The effect of initial SM on the change of SM is explored further in the next section.

Figure 7 presents the respective correlations between SM and T_{soil} at four different depths, representing four layers. According to Pearson's correlation coefficient, SM between two consecutive depths is closely related, e.g., for Day 1 of Experiment A, the top SM (at 2.5 cm depth or the near-surface 0–5 cm layer) had the strongest correlation with the adjacent depth SM (at 7.5 cm depth or the second 10–15 cm layer), having a Pearson correlation coefficient of 0.54. These correlation coefficients decreased with depth, reducing to 0.17, −0.23, and –0.56 at layers of 5–10 cm, 10–15 cm, and 15–20 cm, respectively (Fig. 7a). However, the influence of T_{soil} on SM at different depths did not show any consistent correlation trend with depth (Fig. 7).

The LSTM was used to model the temporal SM variation of different layers (80% of the data for training and remaining 20% for testing for each layer), with and without T_{soil} as input for Day 1 of Experiment A (see Fig. 8). The testing results using the remaining 20% of the data showed good accuracy for all four layers, with Root Mean Square Error (RMSE) of 0.20%, 0.17%, 0.14%, and 0.14% VMC for 0–5 cm, 5–10 cm, 10–15 cm, and 15–20 cm layer, respectively (Table 3). Moreover, when T_{soil} was not included as an input variable, the prediction results showed accuracy similar to when T_{soil} was included. The minimal differences observed in Fig. 8 between curves with and without T_{soil} can be attributed to the dominant influence of other variables in the model. This suggests that while T_{soil} is a contributing factor, its impact is less significant than other variables like initial soil moisture content. Therefore, the LSTM model developed for Day 1 of Experiment A demonstrated the capability to accurately predict SM at different layers solely based on initial SM and time at those layers as input parameters, without using T_{soil} as an additional input. Consequently, this model was employed to predict SM across all four layers in Day 2 of Experiment A, with the results depicted in Fig. 9. The RMSE

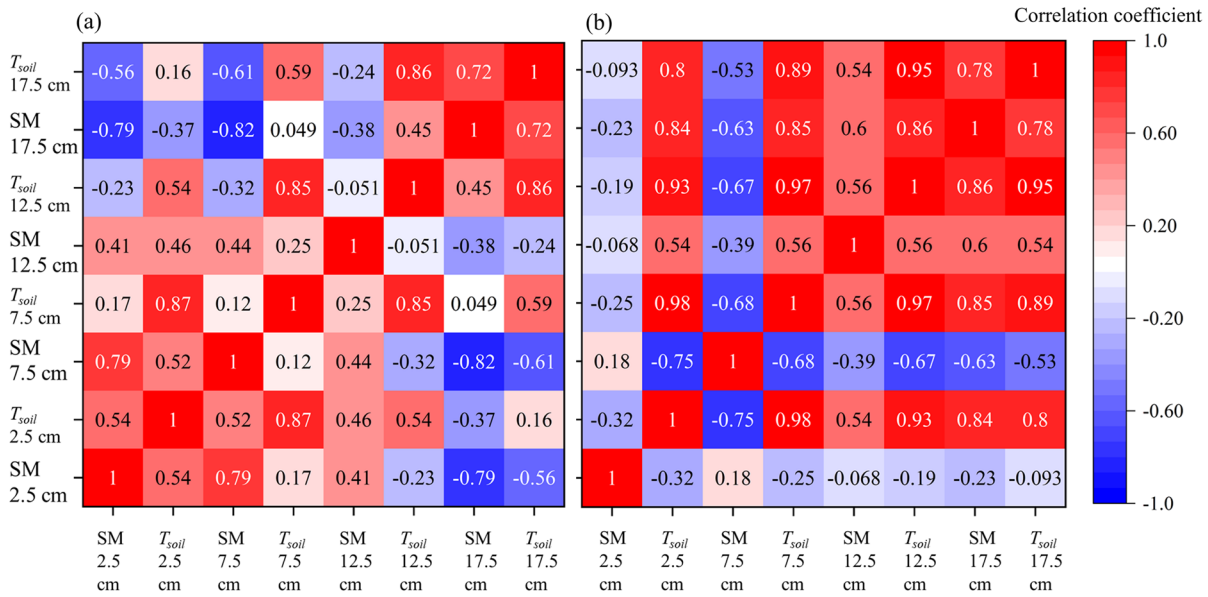

Fig. 7 Heat map of Pearson correlation coefficient between SM and T_{soil} for four different layers (0–5 cm, 5–10 cm, 10–15 cm, and 15–20 cm) calculated for **a** Day 1 of Experiment A and **b** Day 2 of Experiment A. Each coloured square represents the factor used (SM or T_{soil}) at one of 4 different layers, with the colour representing the correlation coefficient results. Layer midpoints are indicated on the axes

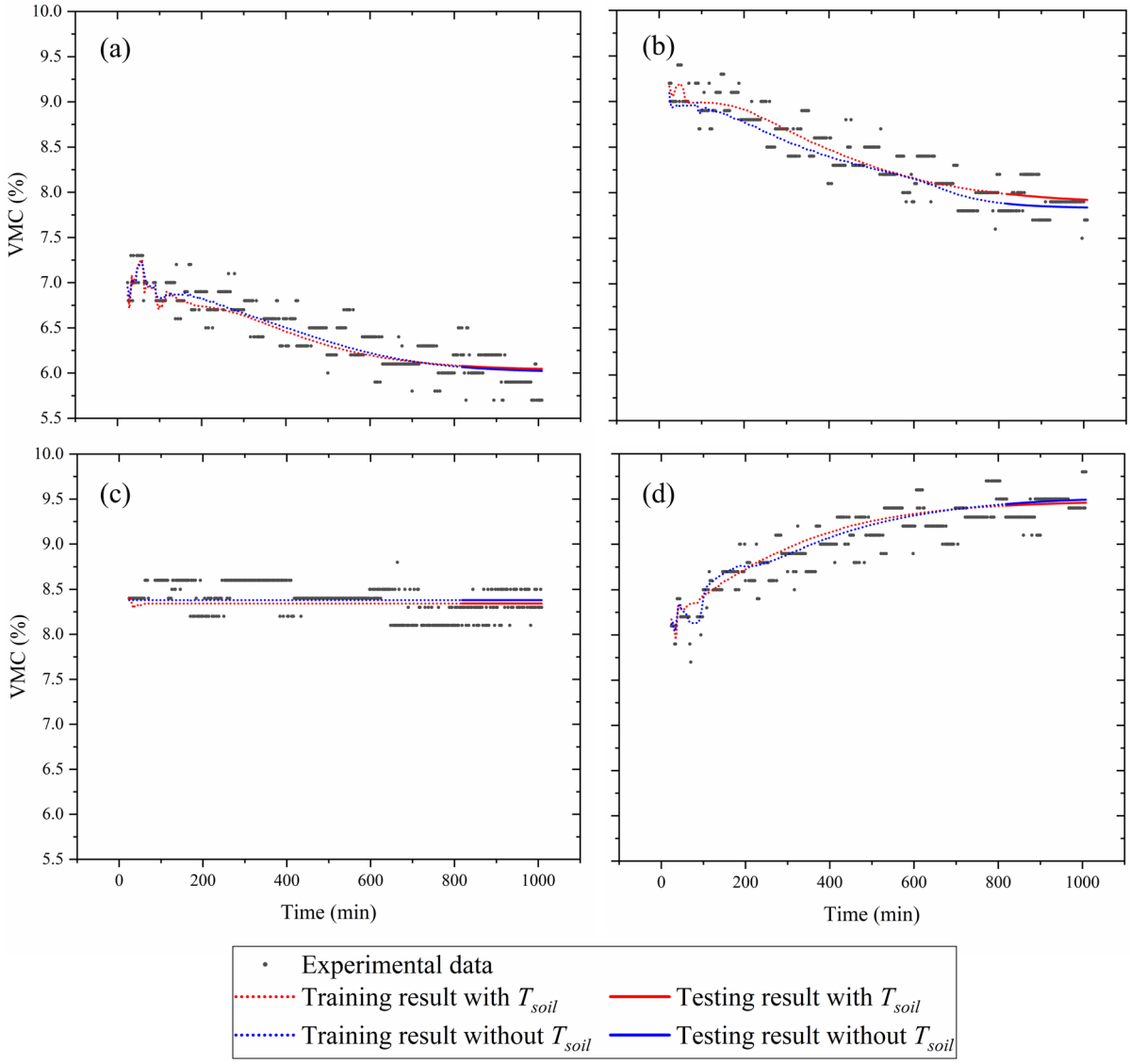

Fig. 8 The experimental and predicted SM results of the LSTM model for training (80% of the data) and testing (remaining 20% of the data) with and without using T_{soil} data for Day 1 of Experiment A for: **a** the near-surface 0–5 cm layer, **b** 5–10 cm layer, **c** 10–15 cm layer, and (d) for the deepest 15–20 cm layer

Table 3 RMSE results (in VMC) of the LSTM model for training (80% of the data) and testing (remaining 20% of the data) with and without using T_{soil} data for Day 1 of Experiment A at all four different layers

Layer	0–5 cm		5–10 cm		10–15 cm		15–20 cm	
	With T_{soil}	Without T_{soil}	With T_{soil}	Without T_{soil}	With T_{soil}	Without T_{soil}	With T_{soil}	Without T_{soil}
Training	0.17	0.16	0.18	0.19	0.19	0.18	0.19	0.17
Testing	0.21	0.20	0.17	0.17	0.14	0.15	0.14	0.14

results indicate that the LSTM model exhibits high accuracy in predicting the SM of all layers for Day 1 of Experiment A, with RMSEs ranging from 0.12% to 0.17% VMC. Furthermore, upon analysing the prediction results, it was observed that a "wave pattern" emerged at the beginning, reflecting the initial behaviour or trend in the predicted SM values. This observation can be attributed to the LSTM model's attempt to replicate the wave pattern based on the initial input received during the prediction process, a sign of initial overfitting in the LSTM model.

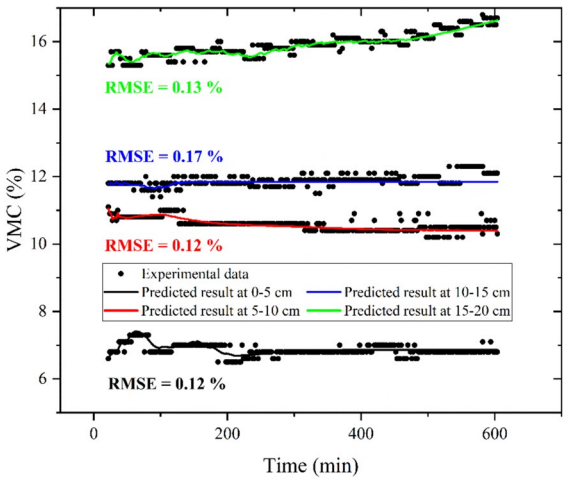

Fig. 9 The experimental and predicted SM results of the LSTM model without using T_{soil} data for Day 2 of Experiment A at all four different layers

The influence of T_{soil} was also evaluated by calculating variable importance (*VI*), as shown in Fig. 10 for different layers. In general, the greatest impact of T_{soil} on SM was at the near-surface layer (0–5 cm) and the deepest layer (15–20 cm), and the smallest effect on SM at the near-deepest layer (10–15 cm). Obtaining T_{soil} at different depths at the construction site can be challenging. As this data did not improve the prediction model's accuracy substantially, soil temperature was not considered in further research steps.

3.2 The Effect of Ambient Conditions on Temporal Variation of Soil Moisture at Different Layers in the Outdoor Environment

For the outdoor environment (Day 1 and Day 2 of Experiment B), the effect of ambient conditions (i.e. *Rn*) on the temporal variation of SM was studied. Figure 11 shows the change of Rn over time. Figure 12 indicates its importance using the *VI* equation on the temporal variation of SM for the three different layers (0–5 cm, 5–10 cm, and 10–15 cm layer) monitored for Experiment B.

Rn was generally found to affect SM variation in all three layers. The results show strong evidence of the *Rn* effect in the near-surface (0–5 cm) layer, where *Rn* influences SM changes with a *VI* of 0.32 for Day 1 of Experiment B and a *VI* of 0.36 for Day 2 of Experiment B. SM was less affected by *Rn* as depth increased, with a *VI* of only 0.15 and 0.26 in the second layer and 0.08 and 0.17 in the third layer

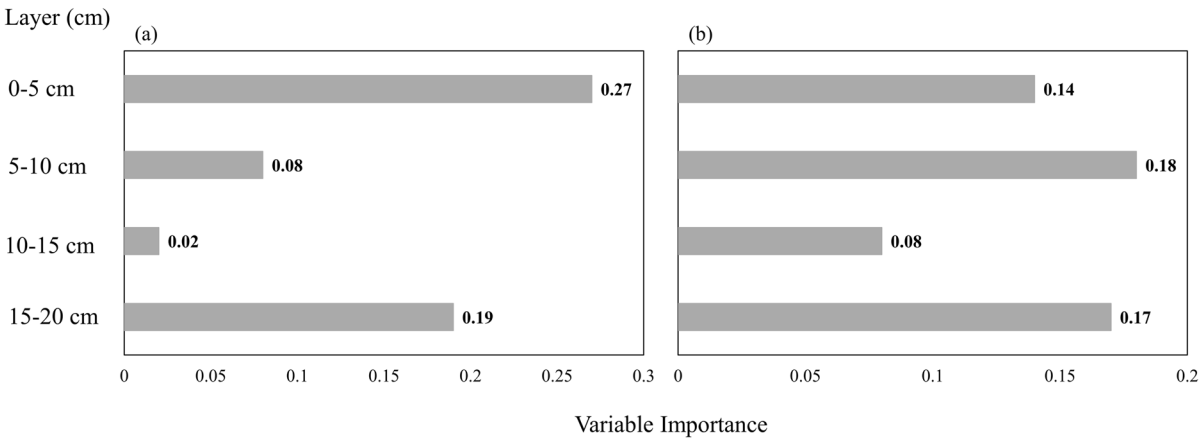

Fig. 10 The variable importance of T_{soil} on SM variation at the four different layers (0–5 cm, 5–10 cm, 10–15 cm, and 15–20 cm) for **a** Day 1 of Experiment A and **b** Day 2 of Experiment A. Layers are indicated on the axes

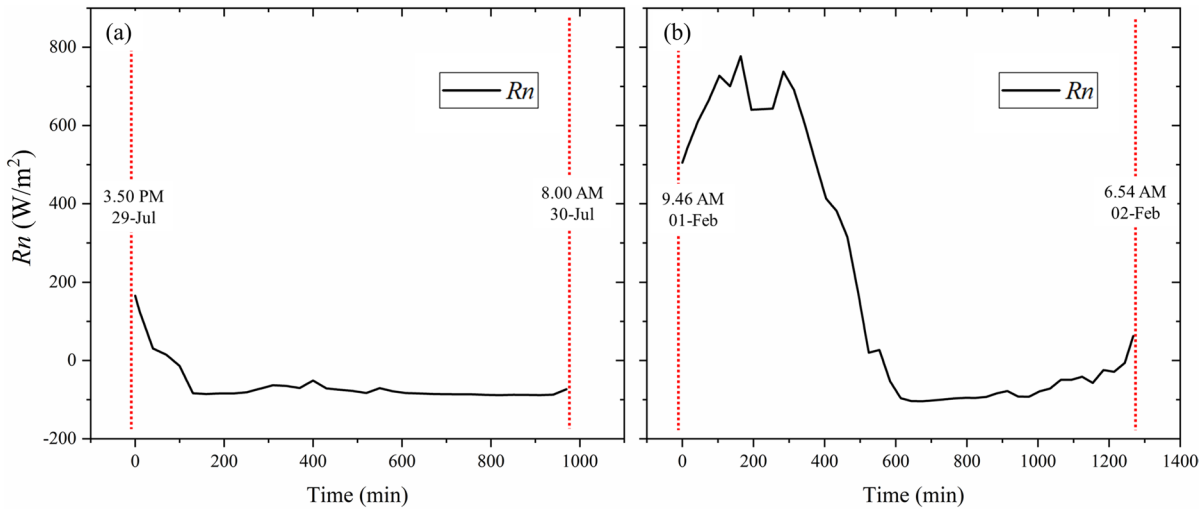

Fig. 11 Variation of *Rn* over time in **a** Day 1 of Experiment B and **b** Day 2 of Experiment B. The red dotted lines indicate the start and end times of recorded data

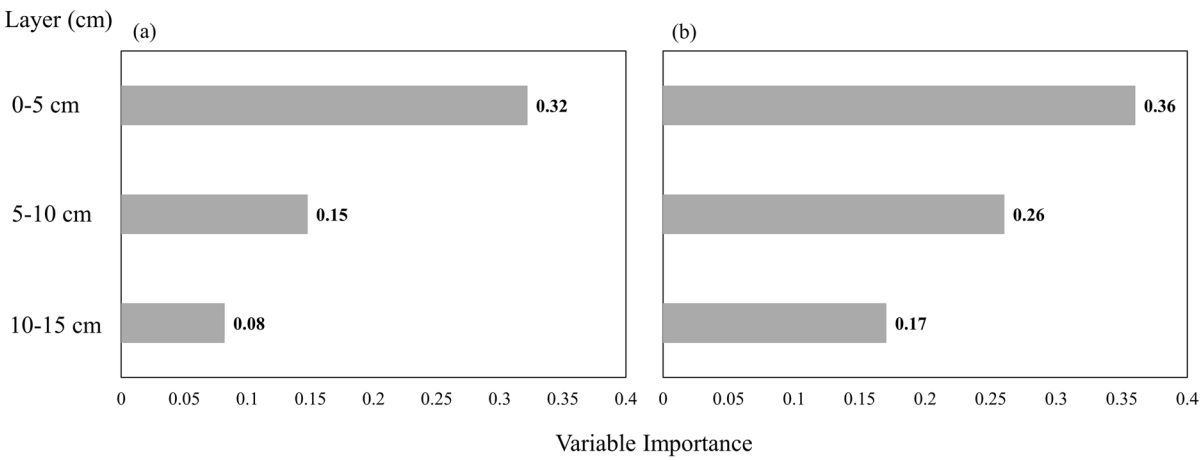

Fig. 12 The variable importance of *Rn* on SM variation at three different layers (0–5 cm, 5–10 cm, and 10–15 cm) calculated for **a** Day 1 of Experiment B and **b** Day 2 of Experiment B. Layers are indicated on the axes

for Experiment B, respectively (Fig. 11). The above two analyses show that the near-surface layer (at 0–5 cm) was most affected by the change in environmental conditions (temperature for indoor and *Rn* for outdoor).

Experiment A, conducted indoors, were designed to isolate soil moisture (SM) variations from rapid ambient changes, thus enabling a controlled analysis of SM dynamics. Comparatively, Experiment B, conducted outdoors, incorporated the net radiation (Rn) variable to study its influence on SM under natural environmental conditions. This distinction is crucial for understanding the model's robustness across controlled and natural settings.

3.3 The Effect of Compaction on Soil Moisture Variation at the Top Layer

Since gravimetric moisture content (GMC) is independent of material density, it is commonly used by field practitioners. Hence, the model development and subsequent analysis emphasised GMC prediction.

In this section, the near-surface layer data from Experiment B were used, and the VMC was converted to GMC using density information from NDG. The near-surface initial GMCs for Day 1 and Day 2 of Experiment B were 10.2% and 3.2% GMC, respectively. These two experiments provided an opportunity to study the effect of temporal variation of SM when the initial SM was in very wet and very dry conditions. The LSTM model was trained and tested using four input parameters, including time, initial GMC, ρ_b, and *Rn*.

Day 2 of Experiment B utilised 80% of the data for training and developing the LSTM model, while the remaining 20% was used for testing. The model demonstrated an accuracy of 0.12% GMC in predicting SM for the last part of Day 2 of Experiment B. Consequently, the same model was employed to predict SM for the near-surface layer in Day 1 of Experiment B, achieving an exceptionally high accuracy of 0.05% GMC (Fig. 13). Figure 14 illustrates the influence of all factors on SM variation. In general, the initial SM conditions ($VI = 0.34$) had the greatest impact on the changes in SM, followed by ρ_b ($VI = 0.29$) and time ($VI = 0.27$), while *Rn* ($VI = 0.10$) had the least impact.

3.4 Prediction of the Temporal Variation of Soil Moisture at Different Layers Using Long Short-Term Memory Model

This section examines a condition where the GMC is not uniform at different depths and when the density data of the interested depths are known. The LSTM model was developed to predict the temporal variation of SM across layers, using just *Rn* and the initial GMC at different layers as inputs. The first 80% of the data from the near-surface layer (0–5 cm) was used for training the model, which achieved a high accuracy compared to experimental data, with an RMSE of 0.12% GMC. Subsequently, the developed model was tested using the remaining 20% of the first near-surface layer (0–5 cm) data, as well as the entire second (5–10 cm) and third layer (10–15 cm) data (Fig. 15). There is strong agreement between LSTM model predictions and the actual measurements at the three different layers, as shown by RMSEs of 0.12%, 0.10%, and 0.09% GMC for 0–5 cm, 5–10 cm, and 10–15 cm layers, respectively. This demonstrates that the LSTM model can be effectively used for practical prediction of SM at deeper layers, utilising the near-surface SM information obtained from the L-band radiometer, *Rn,* and the density data of the interested layers.

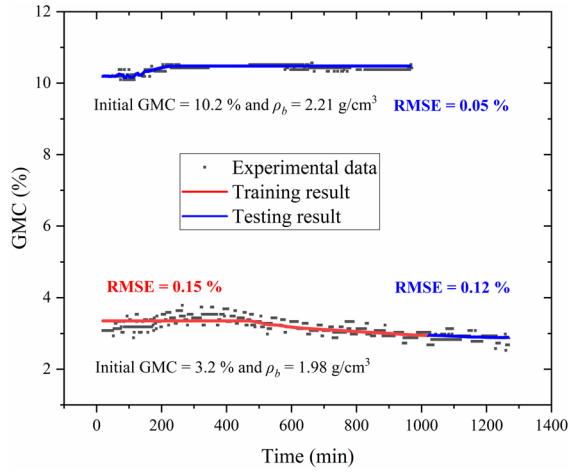

Fig. 13 The experimental and predicted SM results of the LSTM model for the near-surface layer (0–5 cm) for Day 1 of Experiment B (top curve, initial GMC = 10.2%) and Day 2 of Experiment B (bottom curve, initial GMC = 3.2%). The red line in the plot represents the training results obtained using the first 80% of the data from the near-surface layer (0–5 cm) in Day 2 of Experiment B. In contrast, the blue line represents the testing results, including the remaining 20% of the data from the near-surface layer of Day 2 of Experiment B and all the data from Day 1 of Experiment B

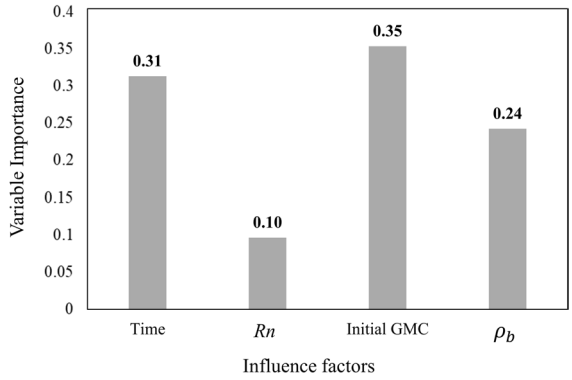

Fig. 14 The variable importance of different factors including time, *Rn*, initial SM content, and ρ_b on SM variation for the near-surface layer (0–5 cm)

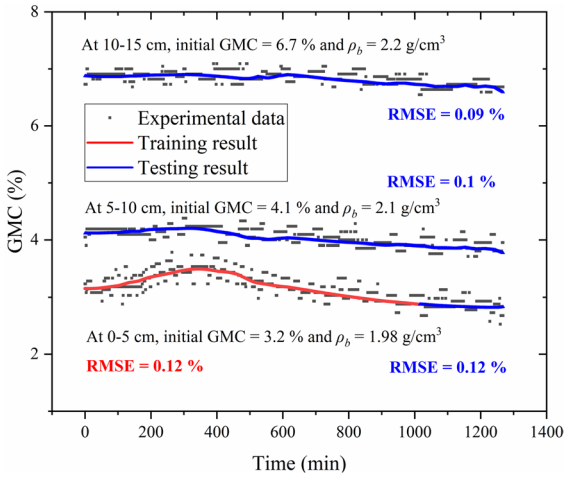

Fig. 15 The observed and predicted SM results of the LSTM model for the three different layers 0–5 cm, 5–10 cm, and 10–15 cm in Day 2 of Experiment B. The red line in the plot represents the training results obtained using the first 80% of the data from the near-surface layer. In contrast, the blue line represents the testing results, which include the remaining 20% of the data from the near-surface layer (0–5 cm), as well as all the data from second (5–10 cm) and third (10–15 cm) layers

3.5 Prediction of the Temporal Variation of Soil Moisture at Different Layers Using the Water Balance Model

Parameters for water balance model were calibrated by minimising the error between the observed data and the predicted results from the water balance model using the first 80% of the data from the first layer of Day 2 of Experiment B. The optimised parameters presented in Table 4. Figure 16 compares the experimental data and the predicted SM results using the water balance model for the calibration and evaluation periods over the three different layers for Day 2 of Experiment B. The RMSE values for the water balance model were higher than those obtained with the LSTM model, indicating lower prediction accuracy. Specifically, the RMSE values for the first, second, and third layers were up to 1.42%, 0.18%, and 0.16% GMC, respectively, which are substantially higher than those obtained with the LSTM model (i.e. 0.12%, 0.1%, and 0.09% GMC for the first, second, and third layers, respectively as indicated in Fig. 15 and Table 5).

The lower prediction accuracy of the water balance model for the first layer can be attributed to several reasons. This study employed the explicit time step formulation in the water balance model. Water balance models often involve various processes that operate at different time scales to ensure accurate simulation. If the time step size is not appropriately chosen, the model can experience numerical instability, resulting in unrealistic outcomes. This study used a time step of 2 min, which may be too large for precise predictions, requiring smaller time steps.

4 Conclusion

In road construction compaction, managing soil moisture (SM) is crucial. This study emphasises the importance of accurately estimating SM's temporal and spatial variation, a key factor in achieving optimal compaction when soil is either too dry or wet and the optimal time for dry back. A Long Short-Term Memory (LSTM) model has been developed to provide real-time SM information for the entire compacted layer, guiding necessary adjustments such as additional watering or natural drying.

Key findings of the study include:

Table 4 Parameters used for three-layer water balance model applied to soil sample 2

Parameters	Layer 1	Layer 2	Layer 3	Unit
Saturated hydraulic conductivity (K_s)	5.02	5.02	5.02	mm/h
Initial water content (VMC)	0.064	0.085	0.148	cm^3/cm^3
Porosity (∅)	0.29	0.25	0.24	cm^3/cm^3
Residual water content (θ_r)	0.022	0.022	0.022	cm^3/cm^3
b	2.103	2.103	2.103	
ψ_a- air entry pressure	−81.416	−81.416	−81.416	mm
Layer depth/thickness	50	50	50	mm
Depth to mid-point of layer	25	75	125	mm

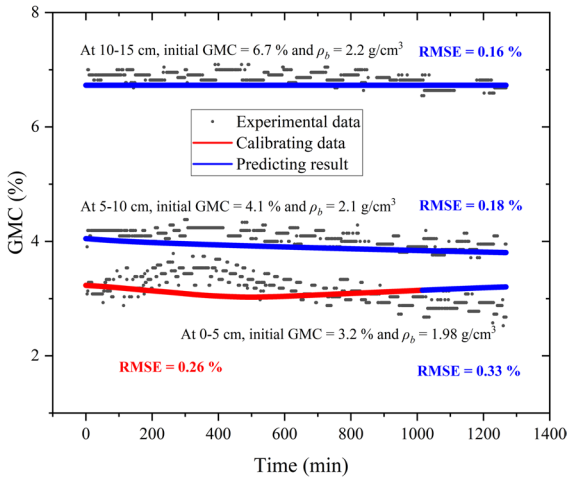

Fig. 16 The observed (black dot), first 80% of the data from the first layer for calibrating (red line), and predicted (blue line) SM results of the water balance model for the three different layers (0–5 cm, 5–10 cm, and 10–15 cm) in Day 2 of Experiment B

Table 5 LSTM vs Water balance model summary

Layer (cm)	LSTM RMSE (% GMC)	Water Balance RMSE (% GMC)
0–5	0.12	1.42
5–10	0.10	0.18
10–15	0.09	0.16

- The LSTM model effectively estimates SM's temporal variation using inputs like net radiation (Rn), initial gravimetric moisture content (GMC), depth, and density.
- It surpasses traditional water balance models in predicting temporal changes in SM, as demonstrated in comparative analyses.
- The model enables accurate predictions of subsurface SM variations from near-surface observations, facilitating efficient compaction.
- Its utility extends to the dry back process in pavement construction, which is crucial for releasing moisture and preventing surface issues such as moisture resurfacing, aggregate punch-in, and premature stiffness loss in pavements, thereby ensuring pavement durability. The dry back process, supported by the LSTM model, is instrumental in maintaining pavement quality and stability.

In conclusion, the LSTM model, possibly in conjunction with L-band radiometer technology (Nguyen 2023; Nguyen et al. 2023; Nguyen et al. 2023), offers a practical, user-friendly approach for on-site SM estimation to achieve Optimum Moisture Content (OMC) before compaction and time for optimal dry back. Its availability as an executable file would simplify implementation, requiring minimal machine learning or coding expertise, thereby making it widely accessible for use in construction projects. Future enhancements to the model, such as the integration of parameters like hydraulic conductivity, as discussed in Richard's equation (Richards 1931), and the adoption of advanced techniques for density measurement (Tophel et al. 2022a, b, c, Tophel et al. 2022a, b, c, Tophel 2023, Tophel et al. 2023) promise to improve its accuracy further and expand its applicability. To get the water needed for the material to reach OMC, in addition to measuring SM at different depths, it is equally important to quantify OMC accurately, which can be challenging due to the material variability in the field. Additionally, it is important to consider that assumptions and simplifications made in the model might not hold for a thick layer, such as a 5 cm thickness, potentially introducing calculation errors. Moreover, the water balance model struggled to capture the variability in SM, particularly the wetting in the middle of the day. This limitation could be attributed to the assumptions made by the water balance model regarding moisture dynamics (i.e. evaporation) not holding or requiring adjustment when dealing with dry moisture conditions. To achieve better results, further work is required to build a more refined water model balance. This can be accomplished by implementing key strategies such as using a smaller time step, a thinner layer, more accurate moisture measurement, and an improved calibration process or using a better model such as Hydrus1D (Simunek 2005). The LSTM model, similar to other machine learning models, should be noted that it depends on high-quality training data. Despite the strong predictive performance demonstrated by the LSTM model, certain limitations must be acknowledged. The model's accuracy is inherently tied to the quality and representativeness of the training data. Its performance may be reduced when applied to soils with significantly different gradations, compaction behaviours, or environmental conditions from those used in training. Additionally, while the model performs well under the

tested field and laboratory conditions, its applicability to other climates or soil types may require further calibration or retraining. Future work should expand the dataset to include more diverse conditions and explore transfer learning techniques to improve the model's generalizability.

Acknowledgements The authors would like to clarify that this manuscript is based on the PhD thesis of the author Thi Mai Nguyen, titled 'Using an L-band radiometer for moisture measurement in road construction,' which was submitted to Monash University in October 2023. The thesis has been appropriately cited in the manuscript. This research work is a part of a research project (Project No IH18.03.2) sponsored by the Smart Pavements Australia Research Collaboration (SPARC) Hub (https://sparchub.org.au) at the Department of Civil Engineering, Monash University, funded by the Australian Research Council (ARC) Industrial Transformation Research Hub (ITRH) Scheme (Project ID: IH180100010). The authors would like to thank the technical staff in the Department of Civil Engineering, Monash University for their help in supporting the experiment. The authors would also like to thank the Smart Pavements Australia Research Collaboration (SPARC) Hub, Construction, Infrastructure, Mining and Concessions (CIMIC), and Engineering, Innovation and Capability (EIC) activities for their financial support, and SPARC Hub members for their moral support throughout the experiments.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. This study was supported by the Australian Research Council (ARC) Industrial Transformation Research Hub (ITRH) Scheme (Grant No. IH180100010).

Data Availability The data that support the findings of this study are available from the corresponding author, AT, upon reasonable request.

Declarations

Competing interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ahmed A, Karimpour A (2025) Machine learning for subgrade moisture in cold regions: random forest and Gaussian mixture models. Geotech Geol Eng 43(2):108

Ahmed AAM, Deo RC, Raj N, Ghahramani A, Feng Q, Yin Z, Yang L (2021) Deep learning forecasts of soil moisture: convolutional neural network and gated recurrent unit models coupled with satellite-derived MODIS, observations and synoptic-scale climate index data. Remote Sens. https://doi.org/10.3390/rs13040554

Ali I, Greifeneder F, Stamenkovic J, Neumann M, Notarnicola C (2015) Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. Remote Sens 7(12):16398–16421

Athmarajah G, Sounthararajah A, Walker JP, Deo R, Kodikara J (2024) An alternative technology using microwaves for dry back process of unbound granular pavements during construction – a review. Transp Geotech. https://doi.org/10.1016/j.trgeo.2024.101245

Brakhasi F, Walker JP, Judge J, Liu P-W, Shen X, Ye N, Wu X, Yeo I-Y, Prajapati R, Kim E, Kerr Y and Jackson T (2024). Multi-layer soil moisture estimation using combined L-and P-band radiometry: an application of machine learning algorithms. IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium**:** 4411–4415.

Briaud J-L (1997). Settlement of bridge approaches, Transportation Research Board.

Broadbent FE (2015). Soil organic matter. Sustain Opt Land Manag.

Brooks RH (1965). Hydraulic properties of porous media, Colorado State University.

Chen DH, Sun R, Yao Z (2013) Impacts of aggregate base on roadway pavement performances. Constr Build Mater 48:1017–1026

Chen L, Chen J, Wang C, Dai Y, Guo R, Huang Q and Mu Y (2021). Modeling of moisture content of subgrade materials in high-speed railway using a deep learning method. Adv Mater Sci Eng **2021**(1).

Clapcott J, Goodwin E and Snelder T (2013). "Predictive models of benthic macroinvertebrate metrics. Prepared for Ministry for the Environment. Retrieved from.

Dec D, Dörner J, Becker-Fazekas O, Horn R (2008) Effect of bulk density on hydraulic properties of homogenized and structured soils. J Soil Sci Plant Nutr 8(1):1–13

Dushmantha A, Zhang R, Gui Y, Zhong J, Gallage C (2025) Deploying machine learning for long-term road pavement moisture prediction: a case study from Queensland, Australia. J Road Eng 5(2):184–201

English N, Weltzin J, Fravolini A, Thomas L, Williams D (2005) The influence of soil texture and vegetation on soil moisture under rainout shelters in a semi-desert grassland. J Arid Environ 63(1):324–343

Fang K, Shen C, Kifer D, Yang X (2017) Prolongation of SMAP to spatiotemporally seamless coverage of continental US using a deep learning neural network. Geophys Res Lett 44(21):11,030-011,039

Filipović N, Brdar S, Mimić G, Marko O, Crnojević V (2022) Regional soil moisture prediction system based on long short-term memory network. Biosyst Eng 213:30–38

Ford T, Harris E, Quiring S (2014a) Estimating root zone soil moisture using near-surface observations from SMOS. Hydrol Earth Syst Sci 18(1):139–154

Ford T, Harris E and Quiring S (2014). Estimating root zone soil moisture using near-surface observations from SMOS. 1foldr Import 2019–10–08 Batch 11.

Gao H-B (2011) Effect of temperature on soil moisture parameters. Adv Water Sci 22(4):484–494

Gao P, Qiu H, Lan Y, Wang W, Chen W, Han X, Lu J (2022) Modeling for the prediction of soil moisture in litchi orchard with deep long short-term memory. Agriculture (Basel) 12(1):25

Geng Q, Yan S, Li Q, Zhang C (2024) Enhancing data-driven soil moisture modeling with physically-guided LSTM networks. Front Forests Glob Change 7:1353011

Grayson RB, Western AW, Chiew FH, Blöschl G (1997) Preferred states in spatial soil moisture patterns: local and nonlocal controls. Water Resour Res 33(12):2897–2908

Gumiere SJ, Camporese M, Botto A, Lafond JA, Paniconi C, Gallichand J, Rousseau AN (2020) Machine learning vs. physics-based modeling for real-time irrigation management. Front Water 2:8

Han H, Choi C, Kim J, Morrison RR, Jung J, Kim HS (2021) Multiple-depth soil moisture estimates using artificial neural network and long short-term memory models. Water 13(18):2584

Hassan M, Gregory P (1999) Water transmission properties as affected by cropping and tillage systems. Pak J Soil Sci 16:29–38

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

Huang Y, Molavi Nojumi M, Hashemian L, Bayat A (2024) Integrating machine learning for improved prediction of temperature and moisture in pavement granular layers. J Test Eval 52(4):2624–2642

Kodikara J, Islam T, Sounthararajah A (2018) Review of soil compaction: history and recent developments. Transp Geotech 17:24–34

Kolassa J, Reichle R, Liu Q, Alemohammad S, Gentine P, Aida K, Asanuma J, Bircher S, Caldwell T, Colliander A (2018) Estimating surface soil moisture from SMAP observations using a neural network technique. Remote Sens Environ 204:43–59

Lim A, Hore-Lacy W, ALDERSON A and Bodin D (2014). Characterisation and performance evaluation of granular bases project: pavement construction report.

Montaldo N, Albertson JD, Mancini M, Kiely G (2001) Robust simulation of root zone soil moisture with assimilation of surface soil moisture data. Water Resour Res 37(12):2889–2900

Montaldo N, Albertson JD, Mancini M (2007) Dynamic calibration with an ensemble Kalman filter based data assimilation approach for root-zone moisture predictions. J Hydrometeorol 8(4):910–921

Nguyen TM, Walker JP, Ye N, Kodikara J (2023) Use of an L-band radiometer for proximal moisture measurement in road construction. Transp Geotech 38:100876

Nguyen TM, Walker JP, Ye N and Kodikara J (2023). "Towards an improved surface roughness parametrization model for soil moisture retrieval in road construction. Transact Geosci Remote Sens.

Nguyen, T. M. (2023). Using an L-band radiometer for moisture measurement in road construction. PhD, Monash University.

OzFlux. (2022). Retrieved 02/06/2022, 2022, from https://www.ozflux.org.au/.

Park S-H, Lee B-Y, Kim M-J, Sang W, Seo MC, Baek J-K, Yang JE, Mo C (2023a) Development of a soil moisture prediction model based on recurrent neural network long short-term memory (RNN-LSTM) in soybean cultivation. Sensors (Basel) 23(4):1976

Park SH, Lee BY, Kim MJ, Sang W, Seo MC, Baek Jk, Yang JE and Mo C (2023). Development of a soil moisture prediction model based on recurrent neural network long short-term memory (RNN-LSTM) in soybean cultivation. Sensors (Basel) **23**(4).

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Rafiq M (1990). Soil variability in agronomic research. Pak J Soil Sci (Pakistan).

Reichle RH, Crow WT, Keppenne CL (2008) An adaptive ensemble Kalman filter for soil moisture data assimilation. Water Resour Res. https://doi.org/10.1029/2007WR006357

Rengasmy PCG (2009). Cation echange capacity, exchangeable cations and solidity. Soil analysis. Australia: an interpretation manual CSIRO publishing.

Richards LA (1931) Capillary conduction of liquids through porous mediums. Physics 1(5):318–333

Salem H (2004) Monitoring and Modeling Subgrade Soil Moisture for Pavement Design and Rehabilitation in Idaho Phase III: Data Collection and Analysis.

Sedgwick P (2012) Pearson's correlation coefficient. BMJ. https://doi.org/10.1136/bmj.e4483

Simunek J (2005). The HYDRUS-1D software package for simulating the movement of water, heat, and multiple solutes in variably saturated medea. HYDRUS Software Series 1.

Stevens. (2022). Reliable Soil Insight. Retrieved 10/10/2020, from https://stevenswater.com/products/hydraprobe/.

Tamboli PM (1961) The influence of bulk density and aggregate size on soil moisture retention.

Tophel A, Walker JP, Dutta TT, Kodikara J (2022a) Theory-guided machine learning to predict density evolution of sand dynamically compacted under Ko condition. Acta Geotech. https://doi.org/10.1007/s11440-021-01431-2

Tophel A, Walker JP, Dutta TT, Kodikara J (2022b) Using a novel instrumented roller to estimate soil dry density during compaction. Trends Const Digital Era Proc ISIC 2022:538–546

Tophel A, Walker JP, Lu Y, Kodikara J (2022c) Proximal sensing of density during soil compaction by instrumented roller. Aust Geomech J 57(3):161–169

Tophel A, Dutta TT, Otsubo M, Kuwano R (2023a) Machine learning models to estimate stress wave velocities of cohesionless soils during triaxial compression influenced by particle characteristics. Soil Dyn Earthq Eng 165:107649

Tophel A, Vogt S, Ramana G (2023c) Investigation of deformation behaviour of uniaxially loaded sand grains using a novel high-resolution imaging apparatus and ensemble machine learning models. Int J Geotech Eng. https://doi.org/10.1080/19386362.2023.2264057

Tophel A, Kodikara J and Walker J (2023). A system and method to measure a deformation of a geomaterial portion due to compaction of the geomaterial portion, WO Patent 2,023,108,190.

Tophel, A. (2023). Proximal estimation of soil density using sensing and modelling. PhD, Monash University.

Verstraeten WW, Veroustraete F, Feyen J (2008) Assessment of evapotranspiration and soil moisture content across different scales of observation. Sensors 8(1):70–117

Wang K, Wang P, Li Z, Cribb M, Sparrow M (2007) A simple method to estimate actual evapotranspiration from a combination of net radiation, vegetation index, and temperature. J Geophys Res Atmos. https://doi.org/10.1029/2006JD008351

Wang Y, Shi L, Hu Y, Hu X, Song W, Wang L (2024) A comprehensive study of deep learning for soil moisture prediction. Hydrol Earth Syst Sci 28(4):917–943

Yinglan A, Wang G, Hu P, Lai X, Xue B, Fang Q (2022) Root-zone soil moisture estimation based on remote sensing data and deep learning. Environ Res 212:113278

Zhang L, Walker GR, Dawes WR (2002) Water balance modelling: concepts and applications. ACIAR Monogr Ser 84:31–47