International Journal of Basic and Applied Sciences, 14 (6) (2025) 242-255

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/wcverd98
Research paper

Comparative Analysis of Machine Learning and Deep Learning Approaches for Hourly Soil Temperature Estimation At Multiple Depths from Meteorological Data in A Semi-Arid Region of Burkina Faso

François Dabilgou, Marcel Bawindsom Kébré *, Soumaila Gandema, Zacharie Koalaga

Laboratoire de Matériaux et Environnement (LAME), Université Joseph KI-ZERBO, Burkina Faso
**Corresponding author E-mail: marcel.kebre@ujkz.bf*

Received: August 30, 2025, Accepted: October 1, 2025, Published: October 12, 2025

Abstract

Accurate estimation of soil temperature is essential for understanding land–atmosphere interactions and managing agricultural systems, par-particularly in semi-arid regions. This study evaluates and compares the performance of multiple modeling approaches: machine learning (Random Forest, XGBoost, LightGBM, SVR, GPR, Decision Tree), and deep learning (MLP-ANN, LSTM, ConvLSTM, Hybrid CNN–LSTM) to predict hourly soil temperature at depths of 10, 20, 30, and 40 cm in a semi-arid land from Burkina Faso. Using high-frequency meteorological data collected, the models were assessed based on their accuracy, temporal generalization, and variable importance. Results show that among ML models, Random Forest demonstrates strong performance up to 30 cm. Deep learning models, particularly ConvLSTM and Hybrid CNN–LSTM, outperform all others across depths, capturing the spatio-temporal variability of subsurface temperature with high fidelity. SHAP analysis reveals that air temperature (Ta) and solar radiation (RAD) dominate at the surface, while dew point tem-temperature (DPT) and vapor pressure deficit (VPD) gain importance in deeper layers. Variables such as rainfall, wind direction, and relative humidity show minimal contribution and can be excluded without sacrificing performance.

Keywords: *Soil Temperature; Subsurface Modeling; Deep Learning; SHAP Analysis; Semi-Arid Climate.*

1. Introduction

Soil temperature is a fundamental variable that regulates energy and water exchanges at the land–atmosphere boundary. It strongly influences processes such as evaporation, soil respiration, microbial activity, and root growth, thereby playing a key role in both hydrological and agricultural systems [1], [2]. In dry tropical semi-arid areas like Burkina Faso, within the Sahel, precise monitoring and forecasting of subsurface soil temperature are crucial for sustainable land management, efficient irrigation scheduling, and climate adaptation planning. With the growing variability of climate, understanding soil thermal dynamics has become increasingly important for analyzing feedbacks in agro-hydro-meteorological systems [3]. Access to detailed, high-resolution soil temperature measurements at different depths is particularly essential for estimating components of the surface energy balance, especially ground heat flux [4]. In addition, as heatwave events have become more frequent in recent years, soil temperature has emerged as a critical parameter for investigating urban heat island effects and radiative processes that shape near-surface microclimates [5].

Soil temperature monitoring relies on diverse approaches, from in situ sensors to satellite-based remote sensing [6]. Ground measurements, using thermocouples or advanced probes, provide reliable high-frequency data but remain limited in the Sahel due to cost and sparse networks, with observations often confined to research trials. Remote sensing products such as MODIS, SMOS, or SMAP enable large-scale retrieval of land surface and subsoil temperatures but suffer from low spatial resolution and limited accuracy over heterogeneous semi-arid terrains, especially without dense ground validation. These constraints highlight the need for numerical estimation methods to generate continuous and representative soil temperature data [7]. Physically based models, such as the one-dimensional heat conduction equation or coupled heat–water transfer models (e.g., HYDRUS), offer accurate simulations but require numerous soil parameters that are rarely available in data-scarce contexts [8], [9]. Consequently, statistical approaches like Multiple Linear Regression (MLR) have long been used to establish empirical links between meteorological variables and soil temperature [10]. More recently, machine learning and deep learning techniques have emerged as powerful alternatives, capable of capturing nonlinear dynamics and improving soil temperature estimation under the challenging conditions of semi-arid tropical regions [6].

Machine learning (ML) methods have gained significant traction in recent years for soil temperature estimation due to their ability to capture nonlinear relationships and interactions among meteorological and environmental variables without relying on complex physical assumptions [3], [7]. As a first example of model performance analysis, [11] investigated the use of machine learning models to estimate

soil temperature in a semi-arid region of Iraqi Kurdistan, focusing on two meteorological stations located in Sulaimani and Dukan. They evaluated Random Forest Regression (RFR), Support Vector Regression (SVR), Gaussian Process Regression (GPR), and Multilayer Perceptron (MLP) models across soil depths of 5 to 100 cm, using explanatory variables such as maximum and minimum air temperature, relative humidity, precipitation, and wind speed. While RFR yielded satisfactory performance, GPR consistently outperformed the other models at several depths, with RMSE values ranging between 1.65 and 1.77 °C. In a second study, [3] compared Decision Trees (DT), Gradient Boosted Trees (GBT/XGBoost), and a hybrid DT–GBT model to predict soil temperature at 5, 10, and 20 cm in a semi-arid continental climate in Divrigi, Turkey. Using air temperature, sunshine duration, and precipitation as inputs, they found that GBT was more accurate at shallow depth (5 cm), while DT or hybrid models performed better at 10 and 20 cm, with accuracy generally increasing with depth. Finally, [12] assessed the prediction of surface soil temperature (0 cm) across diverse tropical and semi-arid zones in China using Random Forest, XGBoost, and M5P (a hybrid model tree). XGBoost achieved the highest performance (average $R^2 = 0.964$ and RMSE = 2.066 °C), closely followed by Random Forest (averaged $R^2 = 0.959$ and RMSE = 2.130°C), while M5P was slightly less accurate (veraged $R^2 = 0.954$ and RMSE = 2.280°C). Tree-based and ensemble ML models (RF, DT, XGBoost) show strong adaptability across climatic zones, with performance varying by depth and local conditions. In tropical semi-arid regions, XGBoost is most effective near the surface (~0–5 cm), while DT or hybrid models perform better at intermediate depths (10–20 cm), offering robust alternatives to traditional statistical methods.

In parallel with machine learning developments, recent literature highlights the growing application of deep learning (DL) techniques for modeling soil temperature across depths and climatic conditions [5], [7], [11]. Among the foundational architectures, multilayer perceptron artificial neural networks (MLP-ANN) remain widely used due to their simplicity and interpretability. For instance, [11] evaluated MLP-ANN alongside SVR, Random Forest Regression (RFR), and Gaussian Process Regression (GPR) for daily soil temperature prediction at depths ranging from 5 to 100 cm in two semi-arid stations in Iraqi Kurdistan. MLP-ANN achieved the best accuracy at 50 cm (RMSE ≈ 2.289 °C), although GPR outperformed it at other depths. For shallow-layer prediction (≤ 7 cm), 1D convolutional neural networks (CNNs) have shown outstanding performance. [13] applied a 1D CNN model in Canada for hourly soil temperature forecasting, demonstrating that CNN significantly outperforms MLP under warm conditions, yielding higher R^2 and lower RMSE. Long Short-Term Memory (LSTM) networks remain the most extensively studied DL architecture for time series-based soil temperature forecasting. [14] Conducted a comparative evaluation of four deep learning models, CNN, LSTM, ConvLSTM, and an attention-enhanced model named ISDNM, using medium fine and very fine soils across diverse climate regions in China, data from the ERA-5 data product for the shallow layer. Their findings demonstrate the superior performance of the ISDNM model, which achieved the highest R^2 value (0.936), the lowest RMSE (9.513), and the lowest MAE (8.320) among all tested models. In comparison, CNN, LSTM, and ConvLSTM yielded lower R^2 values (0.908, 0.916, and 0.923, respectively), along with higher RMSE (11.581, 10.963, 10.472) and MAE (9.743, 9.230, 8.798). These results confirm that integrating attention mechanisms into LSTM architectures, as implemented in ISDNM, substantially enhances predictive accuracy in modeling soil temperature under varying environmental conditions. Furthermore, [15] emphasized in a review the utility of LSTM for modeling soil temperature across depths and seasons. Using SHAP and permutation importance analysis, they showed that 2-meter air temperature is the most influential variable, highlighting LSTM's robustness in seasonal environments. Hybrid CNN–LSTM architectures, which combine spatial feature extraction with sequential memory, are less common but highly promising. [16] proposed a spatio-temporal CNN–LSTM model optimized for 0–7 cm depth, tested across five North American climate zones (humid, continental, subarctic, etc.). The model achieved R^2 between 93.7% and 99.25%, and NRMSE between 1.42% and 3.63%, surpassing RF and SVR. [17] also applied CNN–LSTM to multi-depth soil temperature data, enhancing prediction through multiscale feature extraction and advanced optimization. Similarly, [18] explored ten DL architectures, including feature-attention LSTM (FA-LSTM) and GAN-based LSTM (GAN-LSTM), for soil moisture forecasting, illustrating the added value of attention mechanisms and adversarial training. Their study confirms the modeling strength of LSTM for temporal tasks and demonstrates that FA-LSTM and GAN-LSTM further improve predictive performance. Regarding ConvLSTM, the work of [19] remains a key reference. Using daily ERA5 soil temperature data from northeastern China (0–7 cm), they found that ConvLSTM performed well but was surpassed by a 3D CNN model in their dataset. In contrast, [20] demonstrated the strong performance of ConvLSTM for soil moisture forecasting in China, with $R^2 \approx 0.92$, highlighting its spatio-temporal learning, i.e., joint spatial (e.g., deep) temporal modeling capacity.

[15] evaluated the performance of the LSTM model for soil temperature prediction in a comparative study of this method with other machine learning approaches, such as BPNN (classical neural network) and SVR (Support Vector Regression). The interpretability of the LSTM model was validated through the application of explainable artificial intelligence (XAI) techniques such as SHAP, Permutation Importance (PI), and Partial Dependence Plot (PDP). For this study, 14 predictors (air temperature, precipitation, wind, specific humidity, heat flux, solar and thermal radiation, soil type, topography, etc.) were used. These data were obtained from the LandBench toolbox (ERA5-Land, SoilGrid, MODIS, DEM, etc.), covering the period 1990–2019 for training and 2020 for testing. The study results show that the LSTM model is highly robust and outperforms the BPNN and SVR models (lower RMSE, higher R^2 ~0.70). This performance further increases (R^2 ~0.94) when soil temperature from deeper layers is included. This suggests that the LSTM model is well-suited for comprehensive soil temperature estimation.

Despite their strong performance, most deep learning models for soil temperature estimation remain largely focused on surface or near-surface layers (typically 0–7 cm), with limited exploration of deeper depths. Architectures like ConvLSTM, which offer spatio-temporal learning with lower computational cost, are underutilized. Moreover, attention-based and hybrid models often lack validation in data-scarce tropical contexts. The literature is also dominated by daily or monthly time steps and focused mostly on surface layers; modeling hourly soil temperature at subsurface depths remains underexplored. Given these gaps, we compare in this study multiple modeling approaches to estimate, from in situ meteorological data, hourly soil temperature at various depths (10, 20, 30, and 40 cm) in semi-arid land from Burkina Faso. These include machine learning models (Random Forest (RF), Support Vector Regression (SVR), Gaussian Process Regression (GPR), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Decision Tree(DT), as well as deep learning architectures including Multilayer Perceptron Artificial Neural Network (MLP-ANN), Long Short-Term Memory (LSTM), Convolutional Long Short-Term Memory (ConvLSTM), and a Hybrid Convolutional Neural Network–Long Short-Term Memory model (Hybrid CNN–LSTM).

2. Materials and Methods

2.1. Data collection

The study was conducted in Tanghin, a village located in the rural municipality of Saaba in the Kadiogo Province of Burkina Faso (12°22′5″ N, 1°25′0″ W). The site lies in the Sudano-Sahelian climatic zone, characterized by two distinct seasons: a dry season from November to May and a rainy season from June to October. The average annual rainfall ranges from 600 to 900 mm, and the average temperature varies between 22.6 °C and 35.5 °C [21].
For ground data collection, an 80 cm excavation was made to install sensors and ensure accurate measurements. The soil profile comprised two layers, with the subsurface extending to 30 cm. Meteorological and soil variables were recorded using instruments from Campbell Scientific Inc. (Logan, UT, USA) and METER Group. A ClimaVUE50 autonomous station measured meteorological data. Derived parameters included dew point, wet bulb temperature, actual and saturation vapor pressure, vapor pressure deficit (VPD), and evapotranspiration. Soil monitoring was performed with TEROS21 sensors at 10, 20, 30, and 40 cm, providing soil temperature,

Fig. 1: Field Instrumentation Used for Soil and Meteorological Data Acquisition at the Tanghin Experimental Site (Saaba, Burkina Faso). (Left) The Climavue50 Autonomous Weather Station, mounted on A Metallic Frame and powered by A Solar Panel System, Records Key Atmospheric Variables. (Center) Internal View of the Control Unit Containing the CR350-CELL215 Datalogger and Wiring for Real-Time Data Transmission. (Right) Cross-Sectional Schematic of the Sensor Layout Showing TEROS21 Probes Deployed At 10, 20, 30, and 40 Cm Depths, Each Individually Connected to the Datalogger.

Volumetric water content and matric potential. All variables were logged at 10-minute, hourly, and daily intervals. The data are collected via a CR350-CELL215 datalogger and transmitted to a remote server through 4G for storage and use. Figure 1 shows the station setup.
For the present study, we selected hourly data for the following meteorological variables, denoted by their abbreviations: air temperature (Ta, °C), relative humidity (RH, %), dew point temperature (DPT, °C), wind speed (WS, m/s), wind direction (WD, °), barometric pressure (BP, hPa), precipitation (RAIN, mm), solar radiation (RAD, W/m²), and vapour pressure deficit (VPD, hPa). Soil temperature at various depths is denoted by STx, where x indicates the depth in cm (10, 20, 30, 40).
The data used in this study were collected from 22 September 2023 to 08 March 2025, covering a total period of 534 days. Measurement instrument failures occurred between 01 April 2024 and 14 May 2024, and between 03 and 28 July 2024, resulting in missing data. In total, the dataset comprises 464 days and 11,136 hours (corresponding to 11,136 data rows) for this study. Missing data entries were deleted before analysis.

2.2. Methodology

2.2.1. Machine learning models and deep learning architectures for soil temperature prediction

In contrast to regression analysis, which can suffer significantly from multicollinearity, most machine learning (ML) and deep learning (DL) algorithms, especially tree-based and neural network models, are far less sensitive to highly correlated predictors [22], [23]. These methods do not rely on matrix inversion assumptions and can still produce robust predictive performance even in the presence of collinearity. Therefore, no variable selection was applied before modeling to preserve the full set of meteorological predictors and their potential interactions.

2.2.1.1. Machine learning models used in this study

This study employs several state-of-the-art machine learning (ML) regression models for hourly soil temperature prediction: Random Forest (RF), Support Vector Regression (SVR), Gaussian Process Regression (GPR), Decision Tree (DT), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM). All these models considered fall under the category of supervised learning algorithms, as they rely on labeled input-output data pairs to learn the underlying functional relationships.

- Random Forest (RF) is an ensemble learning technique that combines the predictions of multiple decision trees to improve regression and classification performance. By aggregating several weak learners, RF effectively handles non-linear relationships and reduces overfitting, offering high predictive accuracy [2]. A comprehensive description of RF can be found in [2] and [11].
- Support Vector Regression (SVR) is a kernel-based method derived from Support Vector Machines, designed to solve regression problems by finding a function that deviates from actual targets by a value no greater than a predefined threshold. It performs well with small datasets and non-linear relationships [11]. Mathematical details of SVR are discussed in [11].
- Gaussian Process Regression (GPR) is a non-parametric and probabilistic model used for regression and classification tasks. GPR is well-suited for modeling complex, non-linear systems with limited data, offering predictive uncertainty estimates. Its structure allows for flexible and interpretable modeling. The architecture of GPR is detailed in [11].
- Decision Tree (DT) models use a tree-like structure consisting of nodes (decision points) and branches (decision outcomes), enabling both classification and regression tasks. Each internal node represents a test on a feature, while each leaf node corresponds to an output value. The simplicity and interpretability of DTs make them useful in environmental modeling [3]. [24] Provide a comprehensive review of decision trees, covering their underlying concepts, key algorithms, and practical applications.

- Extreme Gradient Boosting (XGBoost) is an advanced boosting algorithm that builds decision trees sequentially, focusing on correcting the residual errors of previous trees. It has been shown to provide high performance and robustness in many regression applications. A detailed mathematical expression is presented by [12].
- Light Gradient Boosting Machine (LightGBM) is a gradient boosting framework optimized for speed and efficiency, particularly with large datasets. It improves upon XGBoost by using a histogram-based algorithm and leaf-wise tree growth, which enhances accuracy and reduces computational cost. Unlike classical decision trees, LightGBM uses multiple weak learners as nodes within the boosting framework. LightGBM's mathematical formulation is presented in [25] and further discussed by [26] for estimating daily reference evapotranspiration using meteorological data.

These models were selected for their complementary strengths in handling non-linearities, data sparsity, and interactions among meteorological predictors. Readers interested in the detailed mathematical formulations and implementations are encouraged to consult the cited references. The diversity of these models offers distinct strengths for modeling soil temperature, a variable known to be influenced by complex, non-linear interactions between meteorological and soil parameters, with potential temporal autocorrelation and heterogeneity across depths. Using a combination of interpretable (DT, RF) and high-performance (XGBoost, LightGBM, GPR) models allows for a comprehensive evaluation of predictive capabilities, particularly in data-limited contexts where both accuracy and generalization are critical. This approach is especially valuable in semi-arid environments where in-situ soil temperature data are sparse, and robust predictive tools are essential for climate-sensitive applications.

The entire modeling workflow was implemented in Python 3.4 using the scikit-learn library. The dataset was systematically divided into a training set and a test set using an 80/20 split to ensure robust evaluation. The six machine learning regression algorithms use consistent parameters across models (n_estimators = 100, random_state = 42).

2.2.1.2. Deep learning architectures for hourly soil temperature estimation

Four deep learning (DL) architectures were implemented and compared in this study: the Multi-Layer Perceptron Artificial Neural Network (MLP-ANN), the Long Short-Term Memory network (LSTM), a hybrid CNN–LSTM model, and the Convolutional LSTM (ConvLSTM). These models were chosen for their distinct capacities to capture temporal and/or spatial dependencies in environmental data, following methodological frameworks presented in recent literature [5], [7].

- The MLP-ANN serves as a static benchmark model. It consists of fully connected feedforward layers, mapping meteorological inputs directly to soil temperature outputs without incorporating any form of temporal memory. While it is relatively easy to implement and computationally efficient, its inability to account for temporal continuity limits its predictive power when applied to sequential data. Its relevance in soil science modeling remains notable for simple applications, particularly under static or low-frequency conditions, as illustrated in [11].
- The LSTM model extends the capabilities of traditional recurrent neural networks by introducing memory gates that retain and update information across time steps. Its architecture is particularly well-suited for hourly meteorological data, where dependencies span multiple hours or days. LSTM has shown promising results in soil temperature prediction tasks and is widely adopted for its robustness in handling non-stationary time series [1], [2], [26].
- The hybrid CNN–LSTM model combines a convolutional neural network for spatial feature extraction with an LSTM for temporal modeling. This dual architecture enables the network to learn meaningful patterns from multidimensional inputs such as gridded weather data or temporally evolving spatial features. In our study, CNN–LSTM proved effective in capturing both spatial heterogeneity and temporal trends, particularly when using multivariate meteorological inputs. Its structure and applications are discussed in detail by [27] and [28].
- Finally, the ConvLSTM model integrates convolutional operations directly within the LSTM cells, allowing spatial filtering and temporal memory to occur simultaneously. This architecture is especially powerful for spatiotemporal data such as meteorological grids or georeferenced time series. It offers enhanced ability to track evolving thermal patterns in the soil profile by capturing interactions between space and time in a unified framework. The theoretical foundation and performance advantages of ConvLSTM are thoroughly presented in [26] and [29].

Table 1 below presents the set of hyperparameters used for each deep learning architecture. These parameters were manually tuned based on insights from the literature to ensure model robustness while maintaining a computational cost compatible with the processing capacity of the machine used. All LSTM-based models were implemented using Keras (TensorFlow backend) and trained with *MSE* loss and scikit-learn for the MLP.

Table 1: Hyperparameters Used for Each Deep Learning Model Architecture. Parameters Were Manually Optimized Based on the Literature Review and Computational Feasibility

Model	Layers/Structure	Activation(s)	Dropout	Optimizer (LR)	Notes
MLP	2 hidden layers: (100, 50)	ReLU	—	—	Scikit-learn, max_iter = 500
LSTM	LSTM(128, return_sequences = True) → LSTM(64) → Dense(32) → Dense(1)	ReLU, Linear	0.2 (LSTM)	Adam (0.0005)	Sequential Keras model
CNN-LSTM	Conv1D(64, kernel = 3) → MaxPooling1D(2) → Dropout(0.3) → LSTM(100) → LSTM(50) → Dense(50) → Dense(1)	ReLU, Linear	0.3 (CNN+LSTM)	Adam (0.001)	Hybrid CNN-LSTM structure
ConvLSTM	Reshape → ConvLSTM2D(filters = 64, kernel_size = (3,1)) → Flatten → Dense(64) → Dropout(0.2) → Dense(1)	ReLU, Linear	0.2 (Dense)	Adam (0.0005)	ConvLSTM2D input reshaped

2.3. Criteria for model evaluation

To evaluate and compare the performance of the different modeling approaches tested in this study (MLR, ML, and DL), we used a set of classical regression metrics, including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). These indicators quantify both the explanatory power and the predictive accuracy of the models across soil depths.
The formulation of each evaluation metric is presented as follows:

- Mean Absolute Error (MAE)

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N}. \quad (1)$$

- Coefficient of Determination (R^2)

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2}. \quad (2)$$

- Root Mean Square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}}. \quad (3)$$

Where N is the total number of observations, y_i is the observed value at time i, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean of the observed values.
Given that machine learning (ML) and deep learning (DL) methods can incorporate highly correlated or redundant predictors without formal variable selection, a post hoc interpretability technique was applied to understand the internal decision processes. Specifically, SHAP (Shapley Additive exPlanations) analysis was employed to quantify the contribution of each predictor to the model's output [30]. This approach allows for a consistent and theoretically grounded attribution of importance scores to features, even in the presence of complex nonlinearity and multicollinearity. SHAP values were used to interpret model behavior and to compare the relevance of meteorological inputs across soil depths and model types. Such explainability tools have become increasingly common in recent environmental and geospatial applications, including soil temperature modeling [31].

3. Results and Discussion

3.1. Overview of machine learning and deep learning models for soil temperature prediction

This section discusses and interprets the results obtained from the six machine learning (RF, SVR, GPR, XGBoost, LightGBM, DT) and four deep learning (MLP-ANN, LSTM, ConvLSTM, CNN-LSTM) models for hourly soil temperature estimation at different depths (10, 20, 30, and 40 cm) in the Sahelian semi-arid environment of Burkina Faso.

3.1.1. Model performance evaluation across depths

Figures 2 to 4 illustrate density scatter plots comparing observed and predicted soil temperature values at 10 cm and 30 cm depths for both machine learning (ML) and deep learning (DL) models. Only these two depths are shown in the main text to illustrate performance patterns; scatter plots for 20 cm and 40 cm are provided as supplementary figures.
For ML models, Random Forest and XGBoost provided robust predictions at 10 cm with R^2 values close to 0.90 and low error margins (RMSE ≈ 1.5 °C), but performance decreased at 30 cm with a notable drop in R^2 (to ~0.80) and an increase in RMSE. In contrast, DL models such as LSTM and Hybrid CNN–LSTM yielded tighter fits at both depths. For instance, LSTM seems to be better with R^2 values of 0.94 and 0.83 at 10 cm and 30 cm, respectively. Hybrid CNN–LSTM achieved $R^2 = 0.94$ and RMSE < 1.0 °C at 10 cm, outperforming the best overall models.
From a visual standpoint, the shape and orientation of the scatter plots relative to the 1:1 diagonal provide additional insights into model reliability. For Random Forest, XGBoost, and LightGBM, the predicted values align tightly along the diagonal, particularly at 10 cm depth (Figure 2), indicating accurate predictions across the full temperature range. However, a slight widening of the prediction cloud and increased dispersion is observed at 30 cm (Figure 3), reflecting a moderate decline in accuracy as the depth increases. In contrast, SVR, Decision Tree, and especially GPR display more scattered and asymmetric patterns, with systematic underestimation or overestimation at the temperature extremes. The bias becomes more pronounced at 30 cm, where predictions deviate significantly from the diagonal and the point cloud flattens, confirming the poor generalization of these models for subsurface layers.
For DL architectures, the scatter plots (Figure 4) exhibit tight, symmetric distributions closely aligned with the 1:1 diagonal, especially for LSTM, ConvLSTM, and Hybrid CNN–LSTM, confirming their strong capacity to replicate both the central tendency and variability of observed soil temperatures, even at greater depths.
These patterns are consistent with the overall model comparison summarized in the radar plots (Figure 5), where performance metrics, coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE), are jointly visualized for both machine learning (ML) and deep learning (DL) models across all soil depths. The radar layout offers an integrated overview of each model's performance by depth and by metric.
ML models such as Random Forest, XGBoost, and LightGBM showed solid performance in the upper layers (ST10–ST20), but their accuracy degraded with increasing depth. Indeed, in detail, model performance ranking across soil depths shows consistent superiority of XGBoost, followed by LightGBM and Random Forest. At 10 cm depth (ST10), XGBoost achieved the highest coefficient of determination ($R^2 = 0.90$) with a RMSE of 1.04 °C, closely followed by LightGBM ($R^2 = 0.89$; RMSE = 1.10 °C) and Random Forest ($R^2 = 0.88$; RMSE

= 1.08 °C). Similar trends were observed at 20 cm (ST20), where XGBoost again outperformed other models (R^2 = 0.88; RMSE = 1.15 °C), ahead of LightGBM (R^2 = 0.86) and Random Forest (R^2 = 0.85). At deeper layers (ST30 and ST40), XGBoost maintained its leading position with R^2 values of 0.85 and 0.83, respectively, and RMSEs of 1.30 °C and 1.35 °C.

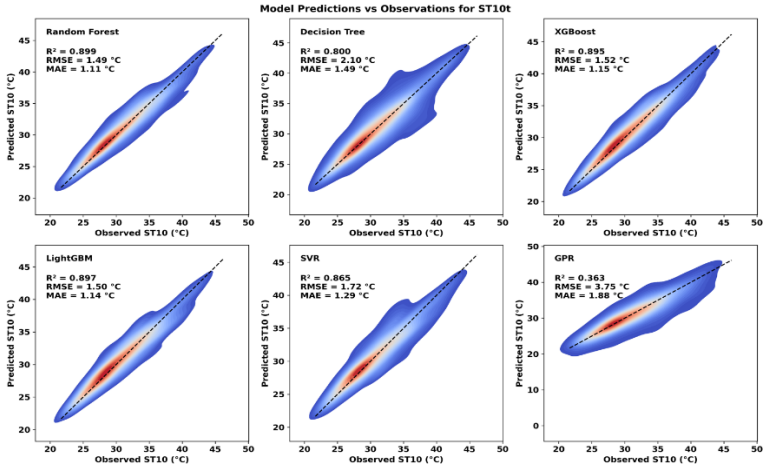

Fig. 2: Predicted Versus Observed Hourly Soil Temperature at 10 Cm Depth Using Machine Learning (ML) Models. Performance Metrics (R^2, RMSE, MAE) Are Shown for Each Depth-Specific Subfigure.

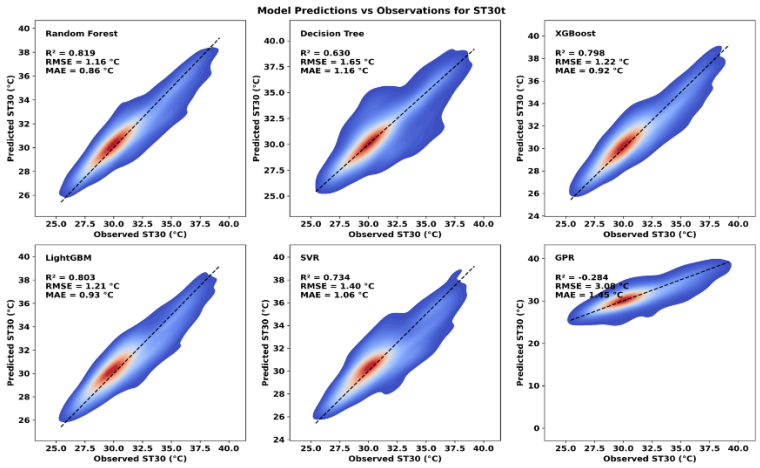

Fig. 3: Predicted Versus Observed Hourly Soil Temperature at 30 Cm Depth Using Machine Learning (ML) Models.

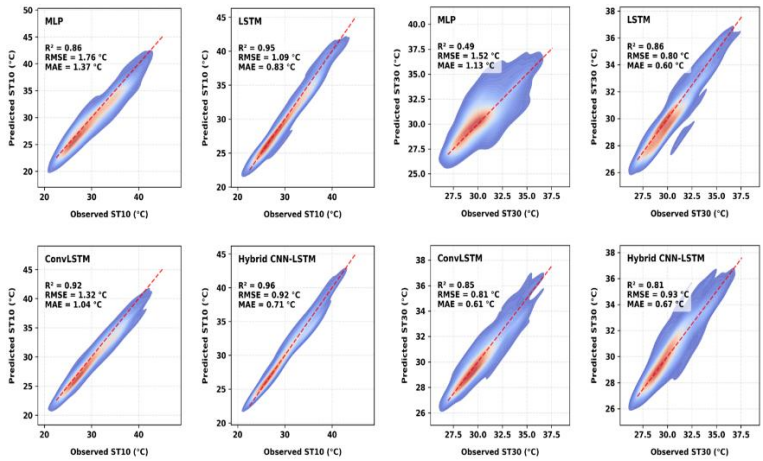

Fig. 4: Predicted Versus Observed Hourly Soil Temperature at 10 Cm and 30 Cm Depths Using Deep Learning (DL) Models. Performance Metrics (R^2, RMSE, MAE) Are Shown for Each Depth-Specific Subfigure.

Across all depths, Decision Tree, SVR, and GPR showed comparatively lower performance, with GPR consistently ranking last (R^2 ranging from 0.67 at ST10 to 0.55 at ST40). These results highlight the robustness of ensemble learning methods, particularly gradient boosting algorithms, in modeling hourly soil temperature dynamics across multiple depths.

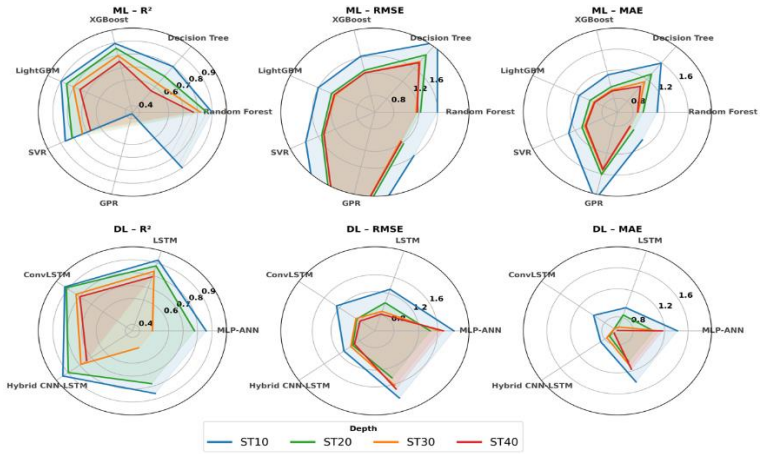

Fig. 5: Radar Plots Comparing Model Performance Across Four Soil Depths (ST10–ST40) For Machine Learning (Top Row) and Deep Learning (Bottom Row) Approaches. Each Plot Summarizes A Specific Metric: Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). Colored Lines Indicate the Model Performance at Each Depth.

Overall, deep learning models show consistent and robust performance across all soil depths, as summarized in the radar plots (Figure 5). At 10 and 20 cm depths, the Hybrid CNN–LSTM slightly outperforms the other models with R^2 values above 0.91 and RMSE below 0.87 °C. From 30 cm downward, ConvLSTM becomes the most accurate model, achieving the highest R^2 (up to 0.918 at 40 cm) and the lowest RMSE (0.610 °C). In contrast, the MLP-ANN, lacking temporal modeling capability, consistently ranks lowest in terms of performance. Conversely, temporal sequence learning in LSTM-based models enabled deep learning approaches to effectively capture the thermal dynamics across all soil layers. These results confirm that hybrid and recurrent DL architectures, such as Hybrid CNN–LSTM and ConvLSTM, are particularly well suited for subsurface soil temperature estimation. Their superior performance, especially at deeper layers, demonstrates their ability to model the spatio-temporal complexity of soil processes. These findings support the adoption of such architectures in environmental modeling under semi-arid conditions like those in Burkina Faso.

Our findings are in line with previous studies reporting the superiority of LSTM-based models for surface soil temperature prediction. For instance, [1] showed that LSTM outperformed other models at 10 cm depth, achieving an R^2 of 0.963 and RMSE of 2.512°C, while XGBoost was the best ML regressor with an R^2 of 0.959 and RMSE of 2.628°C. Similarly, earlier studies such as [3], [11], [12], [14], and [16] emphasized the strong predictive capabilities of DL models, especially LSTM architectures, for soil temperature estimation. Nevertheless, most of these studies focus solely on the topsoil (e.g., 5 or 10 cm), whereas our approach explicitly differentiates performance across multiple depths (10, 20, 30, and 40 cm). Therefore, the general decline in prediction accuracy with increasing depth observed in our study echoes the patterns reported in the literature, reinforcing the importance of depth-specific modeling in subsurface thermal prediction.

3.1.2. Distributional comparison between predicted and observed soil temperatures

To further investigate model behavior, we present in Figures 6 and 7 violin plots comparing the distribution of predicted and observed soil temperature values, which provide insights into each model's ability to reproduce the variability and range of the measured data.

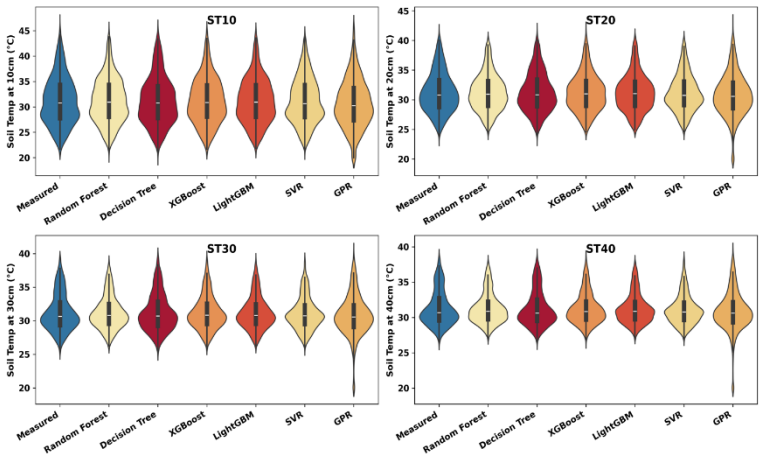

Fig. 6: Violin Plots Comparing Observed and Predicted Hourly Soil Temperature Distributions at 10, 20, 30, and 40 Cm Depths Using Machine Learning Models.

For ML models, at 10 cm, models such as Random Forest, XGBoost, and LightGBM closely replicate the observed distribution (Figure 6), with similar medians and spread. However, as depth increases, discrepancies emerge more clearly: at 30 cm and 40 cm, predicted distributions become narrower and less symmetric, especially for SVR and GPR, which tend to underestimate the temperature range and fail to capture the observed variability. Overall, the most accurate models (Random Forest, XGBoost, LightGBM) maintain a reasonable alignment with the shape and central tendency of the observed data across depths, while less robust models (SVR, GPR) deviate substantially, reflecting their limited ability to reproduce the natural dispersion of soil temperatures under varying environmental conditions.

In contrast to the machine learning models, the deep learning models show a stronger ability to reproduce the full distribution of measured soil temperatures across all depths (Figure 7). At 10 cm, LSTM, ConvLSTM, and Hybrid CNN–LSTM exhibit distributions that closely match the measured data in terms of median, spread, and symmetry, similar to the best-performing ML models (Random Forest, XGBoost, LightGBM). However, unlike ML, this consistency is maintained even at deeper layers (30 cm and 40 cm), where DL models continue to

approximate the shape and dispersion of the observed temperatures. Notably, MLP-ANN shows clear deviation, with narrower and slightly biased distributions, reinforcing the limitations highlighted earlier. These results confirm that while both ML and DL models can perform well near the surface, deep learning architectures provide more robust generalization with depth, better capturing the variability and dynamics of subsurface thermal regimes in semi-arid conditions.

To further interpret model behavior and identify the most influential meteorological predictors, we conducted SHAP (Shapley Additive exPlanations) analyses for the machine and deep learning models. SVR and GPR were excluded from this analysis due to their consistently poor performance across all depths, as previously discussed.

3.2. Depth-specific SHAP analysis of meteorological influences on soil temperature dynamics

3.2.1. SHAP-based feature importance in machine learning models

Figure 8 presents a SHAP (Shapley Additive exPlanations) radar plot of meteorological variable importance for predicting hourly soil temperature at four depths: ST10 (top left), ST20 (top right), ST30 (bottom left), and ST40 (bottom right). Each heatmap displays the

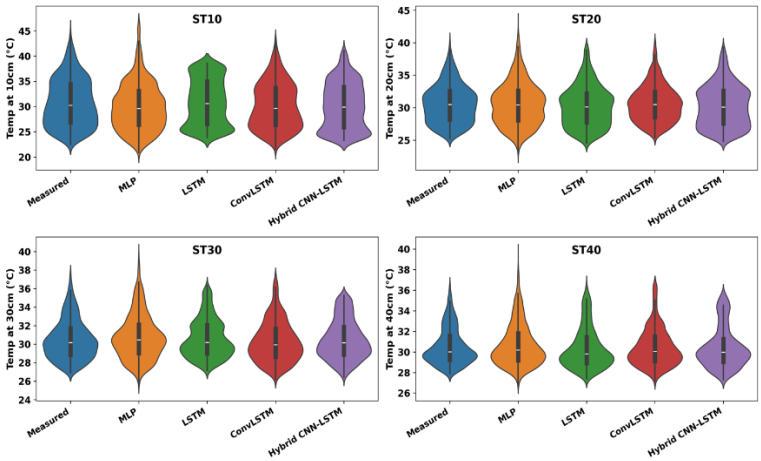

Fig. 7: Violin Plots Comparing Observed and Predicted Hourly Soil Temperature Distributions at 10, 20, 30, and 40 Cm Depths Using Deep Learning Models.

Average SHAP value for nine input variables across four machine learning models: Random Forest, Decision Tree, XGBoost, and LightGBM.

The SHAP radar plot reveals consistent patterns in the contribution of meteorological variables to soil temperature predictions across depths and machine learning models. At 10 cm (ST10), air temperature (Ta) is by far the most influential predictor, with SHAP values exceeding 3.7 across all models. Solar radiation (RAD) follows, with secondary but stable importance (~1.3–1.4). Other variables such as dew point temperature (DPT), vapor pressure deficit (VPD), and wind speed (WS) have modest contributions, while rainfall (RAIN) shows negligible influence across all models (SHAP < 0.03).

As depth increases, the relative importance of Ta decreases, while DPT and VPD become more influential. For instance, at 30 cm (ST30) and 40 cm (ST40), DPT and RAD are consistently among the top contributors, especially in XGBoost and Decision Tree models. This shift reflects the increasing role of subsurface moisture and latent heat in deeper soil layers.

Remarkably, RAIN remains the least relevant variable at all depths, with SHAP values consistently < 0.02. Wind direction (WD) and relative humidity (RH) also show low contribution (typically < 0.2–0.3), particularly in deeper layers. These results suggest that model simplification is possible without significantly degrading performance. Specifically, the following variables can be considered for exclusion: RAIN (negligible contribution at all depths and in all models), WD (wind direction, very low impact, especially beyond 20 cm), and RH (relative humidity, low SHAP values at ST30 and ST40). By focusing on a reduced set of high-impact variables such as Ta, RAD, DPT, and VPD, future models may achieve comparable accuracy with lower computational cost and improved interpretability, especially when deploying in resource-constrained environments.

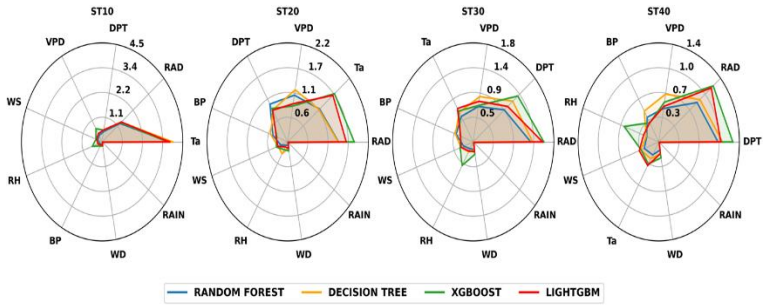

Fig. 8: SHAP Variable Importance Heatmaps for Machine Learning Models Across Four Soil Depths (10 Cm to 40 Cm).

Among the sixteen SHAP violin plots generated for the four models at each depth, four were selected to be presented in the main text for their ability to illustrate contrasting patterns across depths and models: Random Forest at ST10, XGBoost at ST30, Decision Tree at ST40, and LightGBM at ST20. These models represent a range of performance levels and learning strategies, from ensemble-based bagging (Random Forest) to boosting methods (XGBoost, LightGBM) and simpler tree-based learners (Decision Tree). Their inclusion offers a balanced view of both predictive performance and feature interaction across the soil profile.

The selected plots, shown in Figure 9, highlight key transitions in variable importance across depth. At 10 cm depth, Ta and RAD dominate consistently, both in SHAP magnitude and range, confirming their strong contribution to near-surface soil heat dynamics. By 30 cm-depth, DPT and VPD emerge as prominent drivers, and at 40 cm-depth, DPT becomes the primary predictor across all models, with SHAP values broadly distributed and clearly skewed toward positive influence.

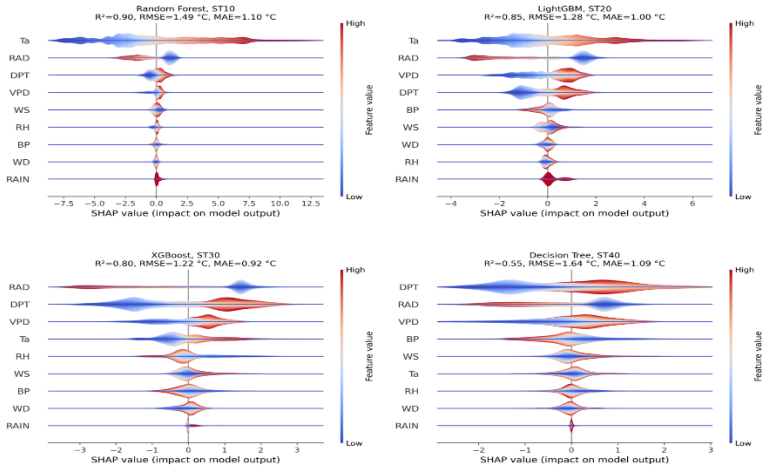

Fig. 9: SHAP Violin Plot for Random Forest at 10 Cm Depth (ST10); Lightgbm at 20 Cm Depth (ST20); Xgboost at 30 Cm Depth (ST30), and Decision Tree at 40 Cm Depth (ST40). Performance Metrics (R^2, RMSE, MAE) Are Shown for Each Depth-Specific Subfigure.

In addition to importance ranking, these violin plots reveal valuable insights into the directionality and nonlinear effects of each variable. For instance, high values of Ta and RAD (shown in red) are consistently associated with increased soil temperature predictions (positive SHAP values), whereas low Ta (in blue) can lead to substantial downward adjustments. At deeper layers, DPT exhibits a strong positive influence for high values and a near-zero or slightly negative impact for low values, confirming its asymmetric effect. VPD and BP also show mixed impacts depending on their range, reflecting more complex interactions. Decision Tree at 40 cm-depth displays the widest spread in SHAP values despite its lower predictive accuracy ($R^2 = 0.55$), suggesting instability in its decision boundaries at depth.
The remaining twelve violin plots, covering all model-depth combinations, are provided in the Supplementary Material to enable comprehensive comparison without overcrowding the main narrative. They confirm the same patterns observed here, further reinforcing the evolving influence of meteorological drivers with soil depth.

3.2.2. SHAP-based feature importance in deep learning models

The SHAP radar plots for deep learning models (Figure 10) reveal a distinct contrast in variable importance patterns when compared to the machine learning results. For MLP-ANN, certain variables clearly dominate, notably Ta, RAD, and VPD, with SHAP values exceeding 3.0 to 4.0 at 10 cm-depth and 40 cm-depth. This behavior mirrors the results seen with ML models and reflects MLP's architecture, which treats input features independently in a fully connected structure.
In contrast, the sequential models (LSTM, ConvLSTM, Hybrid CNN–LSTM) display flattened and low-magnitude SHAP values across most variables and depths, making it difficult to distinguish dominant contributors. This can be explained by the temporal nature of these models: they learn feature influence across time-dependent sequences, where feature importance is distributed dynamically across timesteps. SHAP, in its standard form, may not fully capture these latent interactions, especially when the input is treated as a multidimensional tensor (time × feature). Additionally, feature interactions in LSTM-type models are encoded in memory states, making direct attribution to single input variables less explicit and more diffused.
As a result, although LSTM-based models may achieve high predictive accuracy, their interpretability via standard SHAP heatmaps is more limited, requiring adapted explainability approaches. Nonetheless, the consistent dominance of Ta and DPT in MLP confirms their relevance in driving soil temperature dynamics, particularly at shallow and deep layers.

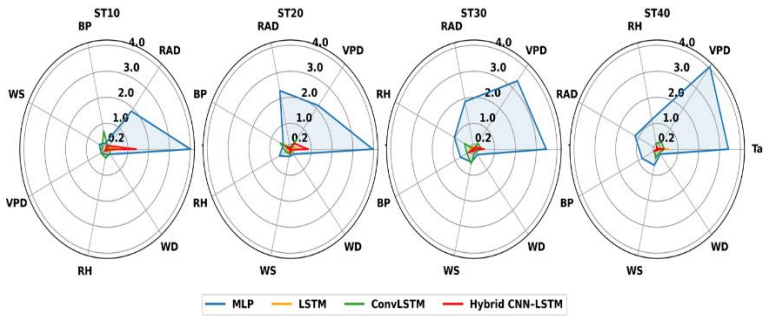

Fig. 10: SHAP Variable Importance Heatmaps for Deep Learning Models Across Four Soil Depths (10 Cm to 40 Cm).

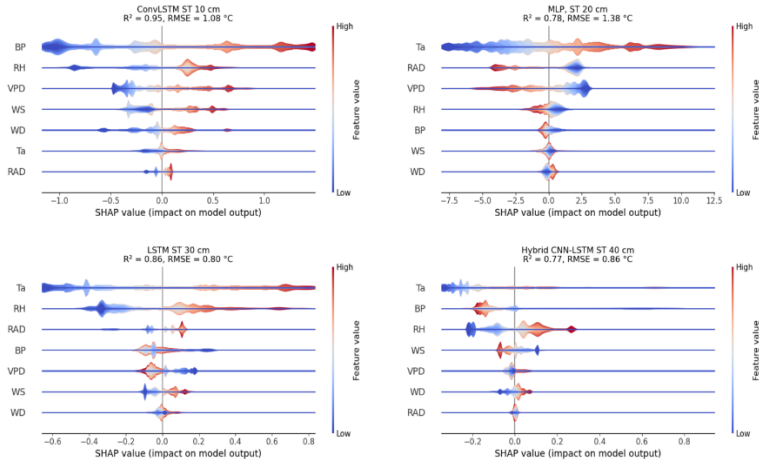

Fig. 11: SHAP Violin Plot for Convlstm at 10 Cm Depth (ST10); MLP-ANN at 20 Cm Depth (ST20); LSTM at 30 Cm Depth (ST30), and Hybrid CNN-LSTM at 40 Cm Depth (ST40). Performance Metrics (R^2, RMSE) Are Shown for Each Depth-Specific Subfigure.

3.3. Discussion on variable importance by depth for ML

The SHAP-based variable importance analysis highlights distinct patterns in how meteorological predictors contribute to soil temperature modeling across depths. As expected, air temperature (Ta) emerges as the most influential variable at 10 cm, followed by solar radiation (RAD). These variables reflect the dominant role of direct energy exchange processes at the soil surface, where temperature dynamics are tightly coupled to short-term atmospheric forcing. However, beyond 20 cm depth, a shift is observed in the hierarchy of explanatory variables. Notably, dew point temperature (DPT) gains increasing importance and often surpasses Ta at 30 cm and 40 cm. For example, DPT shows SHAP values exceeding 1.2 in XGBoost and Decision Tree models at these depths, ranking it among the top two predictors at these depths. This shift likely reflects a change in the governing processes: deeper soil layers are less sensitive to immediate surface fluctuations and more influenced by subsurface moisture content and latent heat transfer. Since DPT captures atmospheric moisture and is directly related to vapor pressure gradients, its increasing influence suggests that moisture-mediated thermal processes (e.g., soil heat capacity modulation, vapor diffusion, and evaporation–condensation cycles) play a larger role with depth. This finding is consistent with established soil physics, where the attenuation of radiative forcing and the dominance of diffusive–conductive heat and moisture fluxes become more pronounced below the surface layer.

Conversely, rainfall (RAIN) exhibits minimal influence across all models and depths, with SHAP values remaining below 0.03. Similarly, wind direction (WD) and, to a lesser extent, relative humidity (RH), show consistently low contributions, particularly at 30 cm and 40 cm. These variables could be excluded from model inputs with little to no loss in predictive power, which may improve model interpretability and computational efficiency. Altogether, these results suggest that a reduced but physically meaningful set of variables, including Ta, RAD, DPT, and VPD, can explain most of the variance in soil temperature across layers. The increasing role of DPT with depth further underscores the need to account for hydrothermal interactions (water-heat dynamics) when modeling subsurface soil temperature, especially under semi-arid conditions where moisture availability is limited and highly variable.

So, in regions where measurement instruments are scarce, such as many parts of the Sahel, focusing on a reduced set of easily measurable parameters, such as air temperature, solar radiation, and vapor pressure deficit, offers a practical and cost-effective approach for accurate soil temperature prediction without compromising model performance.

3.4. Temporal consistency and variability assessment of predictions in soil temperature series

While standard performance metrics and visualizations (e.g., R^2, RMSE, MAE, scatter plots, violin plots) provide useful insights, they often overlook local temporal inconsistencies in time series data. In soil temperature estimation, we therefore complement global metrics with a temporal fidelity analysis to verify the preservation of diurnal and sub-daily variability, ensuring that models capture not only accuracy but also the physical dynamics of soil thermal processes. So, during the simulations for machine and deep learning models, the best-performing model was selected individually for each soil depth based on training and testing metrics. This selected model was then used to generate predictions on the full dataset for final validation and visualization of temporal dynamics. The final predictions presented in Figures 12 and 13 illustrate the temporal evolution of observed and predicted soil temperatures at four depths (10, 20, 30, and 40 cm), using the best-performing model for each depth: Random Forest for machine learning and Hybrid CNN–LSTM for deep learning. Each subplot displays raw values, 24-hour rolling averages, and 95% confidence intervals to facilitate visual assessment of the models' ability to reproduce the natural variability and extrema of the observed time series.

The temporal dynamics of soil temperature predictions reveal distinct patterns across machine learning (ML) and deep learning (DL) models. As shown in the Figures, DL models (Hybrid CNN–LSTM) achieve near-perfect performance ($R^2 \approx 1$), with tightly aligned predictions and observations across all depths. However, this may indicate overfitting during the post-training validation phase, especially as the amplitude and seasonality of fluctuations become visually flattened. By contrast, ML models (Random Forest) exhibit slightly lower R^2 values (0.97–0.99), but still accurately capture the temporal variability of the soil temperature series, including the seasonal transitions. As observed with the MLR models, ML predictions tend to overestimate during the rainy season (JJAS) and underestimate toward the end (OND), suggesting a persistent influence of unmodeled factors such as soil moisture.

This final long-term validation phase is essential not only to assess the overall accuracy of predictions but also to evaluate their robustness over seasonal cycles and across soil depths. Such an analysis ensures that the selected models are capable of capturing both slow seasonal transitions and short-term thermal dynamics in realistic conditions.

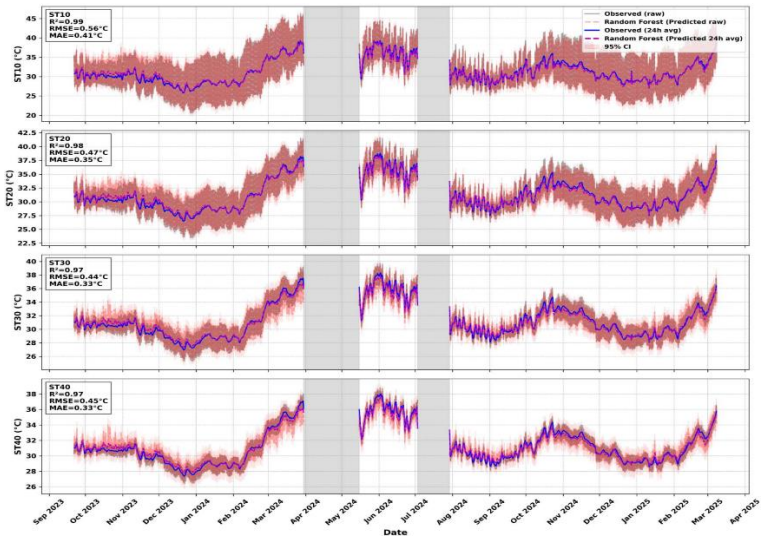

Fig. 12: Temporal Variability of Observed and Random Forest–Predicted Soil Temperatures At 10–40 Cm Depths, with 24-Hour Rolling Means and 95% Confidence Intervals. Validation Performance Metrics (R^2, RMSE, MAE) Are Shown for Each Depth-Specific Subfigure.

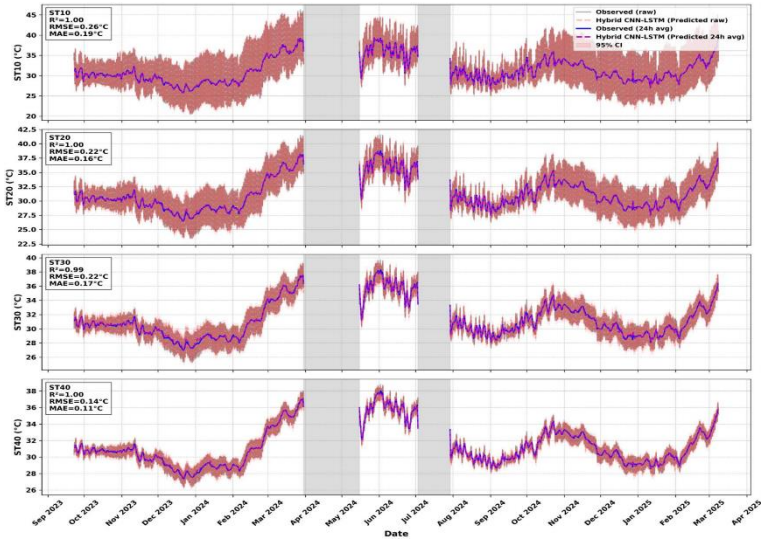

Fig. 13: Temporal Variability of Observed and Hybrid CNN–LSTM–Predicted Soil Temperatures at 10–40 Cm Depths, with 24-Hour Rolling Means and 95% Confidence Intervals. Validation Performance Metrics (R^2, RMSE, MAE) Are Shown for Each Depth-Specific Subfigure.

To further investigate the ability of the models to reproduce fine-scale temporal fluctuations, we conducted a focused comparison over a short time window (15–18 June 2024), which covers a full diurnal cycle over three days. Figures 14 and 15 present the observed and predicted soil temperatures across the four depths for deep learning, machine learning, and multiple linear regression models, respectively, with emphasis on the timing and amplitude of daily extrema.

The short-term validation highlights clear performance differences between the modeling approaches. Deep learning models (Figure 14) exhibit near-perfect agreement with observed data across all depths, accurately reproducing the timing and amplitude of diurnal extrema without noticeable phase lag or amplitude distortion. Machine learning models (Figure 15) also show excellent performance, particularly at 10 and 20 cm depths, where predictions closely track the observed patterns. However, from 30 cm downward, spurious oscillations and increased noise emerge in the predicted series, indicating reduced stability and generalization capacity at greater depths. These findings underscore the importance of model selection based on both global performance metrics and temporal fidelity, especially when targeting physically consistent predictions for environmental applications.

Figure 14 provides a clear visual demonstration of the superior ability of the deep learning (DL) model, specifically the Hybrid CNN–LSTM, to capture the complete diurnal cycle of soil temperature, including its extremes (extrema), across all measured depths. The Hybrid CNN–LSTMs' predictions closely follow the observed data, accurately reproducing both the amplitude and timing of daily peaks and troughs. In contrast, the best-performing machine learning (ML) model, Random Forest (RF), shows limitations in representing these diurnal dynamics (Figure 15). While RF maintains good accuracy at the shallow 10 cm depth, its performance degrades significantly for deeper layers. The introduction of high-frequency noise and the failure to capture the full amplitude of diurnal variations beyond the first layer are evident in the figures.

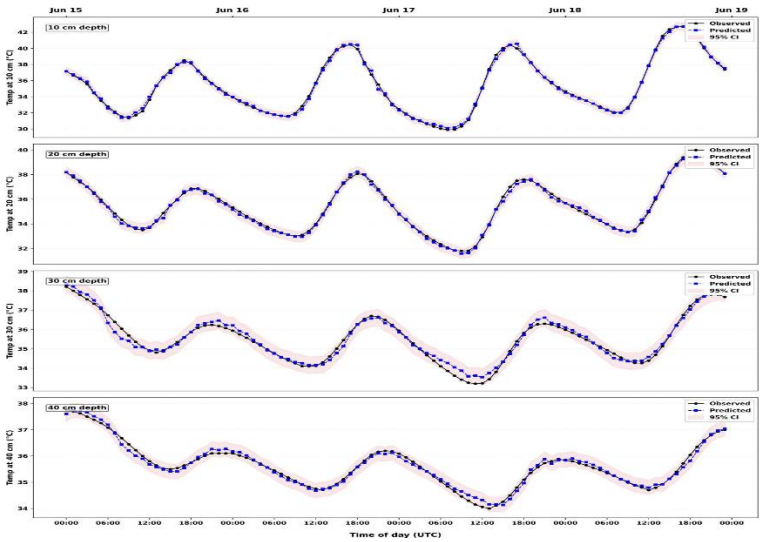

Fig. 14: Short-Term Validation of Deep Learning Model (Hybrid CNN–LSTM) Predictions Versus Observations at 10–40 Cm Depths from 15 to 18 June 2024.

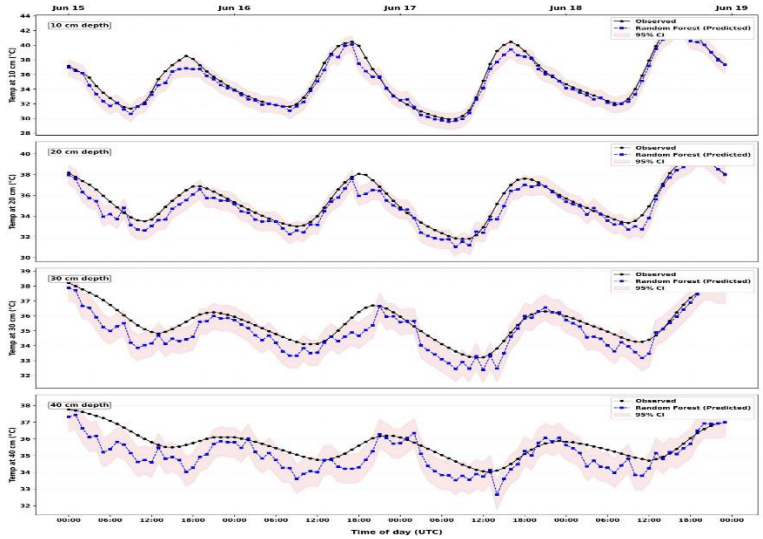

Fig. 15: Short-Term Validation of Machine Learning Model (Random Forest) Predictions Versus Observations at 10–40 Cm Depths from 15 to 18 June 2024.

4. Conclusion

This study presents a comprehensive comparative analysis of machine learning (ML), and deep learning (DL) approaches to estimate hourly soil temperature at different depths of 10, 20, 30, and 40 cm in a semi-arid region of Burkina Faso. Conducted at the Tanghin experimental site, the study highlights the challenges and opportunities associated with modeling subsurface thermal dynamics in low-instrumentation contexts and under high temporal resolution.

Among the ML models tested, Random Forest performs particularly well up to 30 cm depth, maintaining high R^2 values and accurately reproducing temporal variability and seasonal transitions. Beyond this depth, however, ML models show reduced generalization capacity, with increasing noise and oscillations in the predicted series. Deep learning models, particularly ConvLSTM and Hybrid CNN–LSTM, clearly outperform all others across depths. These architectures not only achieve near-perfect alignment with observed values but also generalize more effectively in deeper layers, capturing both the magnitude and timing of soil temperature dynamics. Unlike MLP-ANN or even traditional ML models, LSTM-based models exhibit strong consistency in distributional shape and temporal variation, even under seasonal shifts. However, potential overfitting during short-term validation suggests the need for careful model evaluation beyond accuracy metrics alone.

Interpretability analysis using SHAP values confirms the dominant role of air temperature (Ta) and solar radiation (RAD) near the surface, and a progressive shift toward dew point temperature (DPT) and vapor pressure deficit (VPD) with depth. These findings are consistent with the underlying soil physics, where heat and moisture transport processes such as vapor diffusion and latent heat transfer govern subsurface thermal behavior. Variables like rainfall (RAIN), wind direction (WD), and relative humidity (RH) contribute minimally and may be excluded to simplify models without performance loss.

Altogether, the results support the use of a reduced yet physically meaningful set of predictors (Ta, RAD, DPT, VPD) for subsurface soil temperature modeling. DPT's rising influence with depth also highlights the need to incorporate hydrothermal interactions in future models. For operational use in data-scarce environments like the Sahel, these insights pave the way for more cost-effective and scalable modeling strategies, balancing model complexity with measurement feasibility. Incorporating soil moisture as a predictor remains a critical next step to improve generalization, particularly during the rainy season, when water content and latent heat fluxes dominate thermal processes in the subsurface.

In this present study, the hyperparameters used for the final results were chosen to optimize computational time in a resource-constrained environment and to mitigate overfitting through the specific proportions allocated for the test and validation sets.

Despite its comprehensive scope, this study has some limitations. The models were trained on data from a single semi-arid site, which may reduce transferability to areas with different soil or climatic conditions. The absence of continuous soil moisture measurements, a key driver of subsurface thermal dynamics, likely contributed to reduced performance during the wet season. Moreover, although deep learning models achieved strong accuracy, their high complexity raises concerns about interpretability and potential overfitting. Future research should extend validation across multiple agroecological zones, integrate soil moisture and thermal properties, and explore hybrid physics-informed approaches to improve robustness and transparency. Coupling soil temperature models with crop growth, irrigation, or land surface models could also enhance their practical value for precision agriculture and climate risk monitoring in semi-arid regions. Furthermore, the following research question will be addressed: How can satellite data enhance model transferability in a region where in-situ data is largely unavailable?

Acknowledgement

We gratefully acknowledge the support of the Higher Education Support Project (PAES) in Burkina Faso, funded by the World Bank, for providing a doctoral research equipment grant to Mr. Dabilgou.

References

[1] Mampitiya, L., Rozumbetov, K., Rathnayake, N., Erkudov, V., Esimbetov, A., Arachchi, S., ... & Rathnayake, U. (2024). Artificial intelligence to predict soil temperatures by development of novel model. *Scientific Reports*, *14*(1), 9889. https://doi.org/10.1038/s41598-024-60549-x.

[2] Elsayed, S., Gupta, M., Chaudhary, G., Taneja, S., Gaur, H., Gad, M., ... & Schmidhalter, U. (2023). Interpretation the influence of hydrometeorological variables on soil temperature prediction using the potential of deep learning model. *Knowledge-Based Engineering and Sciences*, *4*(1), 55-77. https://doi.org/10.51526/kbes.2023.4.1.55-77.

[3] Sattari, M. T., Avram, A., Apaydin, H., & Matei, O. (2020). Soil Temperature Estimation with Meteorological Parameters by Using Tree-Based Hybrid Data Mining Models. *Mathematics*, *8*(9), 1407. https://doi.org/10.3390/math8091407.

[4] Hsieh, C. I., Chiu, C. J., Huang, I. H., & Visessri, S. (2023). Estimating Ground Heat Flux from Net Radiation. *Atmosphere*, *14*(12), 1778. https://doi.org/10.3390/atmos14121778.

[5] Edmondson, J. L., Stott, I., Davies, Z. G., Gaston, K. J., & Leake, J. R. (2016). Soil surface temperatures reveal moderation of the urban heat island effect by trees and shrubs. *Scientific Reports*, *6*(1), 33708. https://doi.org/10.1038/srep33708.

[6] Asadzadeh, F., Emami, S., Elbeltagi, A., Akiner, M. E., Rezaverdinejad, V., Taran, F., & Salem, A. (2025). Investigating the impact of meteorological parameters on daily soil temperature changes using machine learning models. *Scientific Reports*, *15*(1), 19988. https://doi.org/10.1038/s41598-025-04605-0.

[7] Taheri, M., Schreiner, H. K., Mohammadian, A., Shirkhani, H., Payeur, P., Imanian, H., & Cobo, J. H. (2023). A review of machine learning approaches to soil temperature estimation. *Sustainability*, *15*(9), 7677. https://doi.org/10.3390/su15097677.

[8] Šimůnek, J., Brunetti, G., Jacques, D., van Genuchten, M. T., & Šejna, M. (2024). Developments and applications of the HYDRUS computer software packages since 2016. *Vadose Zone Journal*, *23*(4), e20310. https://doi.org/10.1002/vzj2.20310.

[9] Zhang, Y., Li, X., Šimůnek, J., Chen, N., Hu, Q., & Shi, H. (2024). Evaluating the effects of different irrigation water sources on soil temperature using HYDRUS (2D/3D) and considering the coupled movement of water and heat. *Soil and Tillage Research*, *244*, 106259. https://doi.org/10.1016/j.still.2024.106259.

[10] Progga, J. F., Khan, M. N. H., & Amın, M. M. (2023). Meteorological Parameters–Soil Temperature Relations in a Sub-Tropical Summer Grassland: Physically-Based and Data-Driven Modeling. *Atatürk Üniversitesi Ziraat Fakültesi Dergisi*, *54*(2), 48-56. https://doi.org/10.5152/AUAF.2023.23126.

[11] Alizamir, M., Ahmed, K. O., Kim, S., Heddam, S., Gorgij, A. D., & Chang, S. W. (2023). Development of a robust daily soil temperature estimation in semi-arid continental climate using meteorological predictors based on computational intelligent paradigms. *PLoS One*, *18*(12), e0293751. https://doi.org/10.1371/journal.pone.0293751.

[12] Alizamir, M., Kisi, O., Ahmed, A. N., Mert, C., Fai, C. M., Kim, S., ... & El-Shafie, A. (2020). Advanced machine learning model for better prediction accuracy of soil temperature at different depths. *PLoS One*, *15*(4), e0231055. https://doi.org/10.1371/journal.pone.0231055.

[13] Farhangmehr, V., Cobo, J. H., Mohammadian, A., Payeur, P., Shirkhani, H., & Imanian, H. (2023). A Convolutional Neural Network Model for Soil Temperature Prediction under Ordinary and Hot Weather Conditions: Comparison with a Multilayer Perceptron Model. *Sustainability*, *15*(10), 7897. https://doi.org/10.3390/su15107897.

[14] Li-Xiaoning, X., Zhu, Y., Li, Q., Zhao, H., Zhu, J., & Zhang, C. (2023). Interpretable spatio-temporal modeling for soil temperature prediction. *Frontiers in Forests and Global Change*, *6*, 1295731. https://doi.org/10.3389/ffgc.2023.1295731.

[15] Geng, Q., Wang, L., & Li, Q. (2024). Soil temperature prediction based on explainable artificial intelligence and LSTM. *Frontiers in Environmental Science*, *12*, 1426942. https://doi.org/10.3389/fenvs.2024.1426942.

[16] Farhangmehr, V., Imanian, H., Mohammadian, A., Cobo, J. H., Shirkhani, H., & Payeur, P. (2025). A spatiotemporal CNN-LSTM deep learning model for predicting soil temperature in diverse large-scale regional climates. *Science of The Total Environment*, *968*, 178901. https://doi.org/10.1016/j.scitotenv.2025.178901.

[17] Mohammed, S., Arshad, S., Ocwa, A., Al-Dalahmeh, M., ALDabbas, A., Alzoubi, M. M., ... & Harsányi, E. (2025). Advanced Attention-Driven Deep Learning Architectures for Multi-Depth Soil Temperature Prediction. *Results in Engineering*, 106508. https://doi.org/10.1016/j.rineng.2025.106508.

[18] Wang, Y., Shi, L., Hu, Y., Hu, X., Song, W., & Wang, L. (2023). A comprehensive study of deep learning for soil moisture prediction. *Hydrology and Earth System Sciences Discussions*, *2023*, 1-38. https://doi.org/10.5194/hess-28-917-2024.

[19] Yu, F., Hao, H., & Li, Q. (2021). An Ensemble 3D Convolutional Neural Network for Spatiotemporal Soil Temperature Forecasting. *Sustainability*, *13*(16), 9174. https://doi.org/10.3390/su13169174.

[20] Huang, F., Zhang, Y., Zhang, Y., Shangguan, W., Li, Q., Li, L., & Jiang, S. (2023). Interpreting Conv-LSTM for spatio-temporal soil moisture prediction in China. *Agriculture*, *13*(5), 971. https://doi.org/10.3390/agriculture13050971.

[21] Ouoba Nebie, B. A., Lawane, A., Hema, C., & Siroux, M. (2025). Energy efficiency in the building sector in Burkina Faso: Literature review, SWOT analysis, and recommendations. *Energies*, 18(11), 2689. https://doi.org/10.3390/en18112689.

[22] Yıldırım, H. (2024). The multicollinearity effect on the performance of machine learning algorithms: Case examples in healthcare modelling. *Asia-Pacific Journal of Educational, Social and Scientific Research*, *12*(3), 68–80. https://doi.org/10.21541/apjess.1371070.

[23] Chan, J. Y.-L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z.-W., & Chen, Y.-L. (2022). Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics*, *10*(8), 1283. https://doi.org/10.3390/math10081283.

[24] Mienye, I. D., & Jere, N. (2024). A survey of decision trees: Concepts, algorithms, and applications. *IEEE access*, *12*, 86716-86727. https://doi.org/10.1109/ACCESS.2024.3416838.

[25] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.

[26] Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., & Zeng, W. (2019). Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agricultural water management*, *225*, 105758. https://doi.org/10.1016/j.agwat.2019.105758.

[27] Airlangga, G.; Bata, J.; Adi Nugroho, O.I.; Lim, B.H.P. (2025). Hybrid CNN-LSTM Model with Custom Activation and Loss Functions for Predicting Fan Actuator States in Smart Greenhouses. *AgriEngineering* , 7, 118. https://doi.org/10.3390/agriengineering7040118.

[28] Qian, L., Wu, L., Zhang, Z., Fan, J., Yu, X., Liu, X., ... & Cui, Y. (2024). A gap filling method for daily evapotranspiration of global flux data sets based on deep learning. *Journal of Hydrology*, *641*, 131787. https://doi.org/10.1016/j.jhydrol.2024.131787.Li, A.; Zhang, W.; Zhang, X.; Chen, G.; Liu, X.; Jiang, A.; Zhou, F.; Peng, H. (2024). A Deep U-Net-ConvLSTM Framework with Hydrodynamic Model for Basin-Scale Hydrodynamic Prediction. *Water*, 16, 625. https://doi.org/10.3390/w16050625.

[29] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *arXiv*. https://arxiv.org/abs/1705.07874.

[30] Douanla, A. Y., Dembélé, A., Ossénatou, M., & Lenouo, A. (2022, February). *Prediction of daily direct solar energy based on XGBoost in Cameroon and key parameter impacts analysis*. In *2022 IEEE Multi-conference on Natural and Engineering Sciences for Sahel's Sustainable Development (MNE3SD)* (pp. 1–7). IEEE. https://doi.org/10.1109/MNE3SD53781.2022.9723309.