


## Article

# GRU–Transformer: A Novel Hybrid Model for Predicting Soil Moisture Content in Root Zones

Wengang Zheng<sup>1,2</sup>, Kai Zheng<sup>1</sup>, Lutao Gao<sup>3</sup>, Lili Zhangzhong<sup>2,4,\*</sup>, Renping Lan<sup>2</sup>, Linlin Xu<sup>5</sup>  
and Jingxin Yu<sup>2,4,\*</sup> 

<sup>1</sup> College of Agricultural Engineering, Shanxi Agricultural University, Jinzhong 030801, China

<sup>2</sup> Intelligent Equipment Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

<sup>3</sup> School of Big Data, Yunnan Agricultural University, Kunming 650201, China

<sup>4</sup> Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

<sup>5</sup> Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

\* Correspondence: zhangzll@nercita.org.cn (L.Z.); yujx@nercita.org.cn (J.Y.)

**Abstract:** The accurate measurement of soil moisture content emerges as a critical parameter within the ambit of agricultural irrigation management, wherein the precise prediction of this variable plays an instrumental role in enhancing the efficiency and conservation of agricultural water resources. This study introduces an innovative, cutting-edge hybrid model that ingeniously integrates Gated Recirculation Unit (GRU) and Transformer technologies, meticulously crafted to amplify the precision and reliability of soil moisture content forecasts. Leveraging meteorological and soil moisture datasets amassed from eight monitoring stations in Hebei Province, China, over the period from 2011 to 2018, this investigation thoroughly assesses the model's efficacy against a diverse array of input variables and forecast durations. This assessment is concurrently contrasted with a range of conventional machine learning and deep learning frameworks. The results demonstrate that (1) the GRU–Transformer model exhibits remarkable superiority across various aspects, particularly in short-term projections (1- to 2-day latency). The model's mean square error (MSE) for a 1-day forecast is notably low at 5.22%, reducing further to a significant 2.71%, while the mean coefficient of determination ( $R^2$ ) reaches a high of 89.92%. Despite a gradual increase in predictive error over extended forecast periods, the model consistently maintains robust performance. Moreover, the model shows exceptional versatility in managing different soil depths, notably excelling in predicting moisture levels at greater depths, thereby surpassing its performance in shallower soils. (2) The model's predictive error inversely correlates with the reduction in parameters. Remarkably, with a streamlined set of just six soil moisture content parameters, the model predicts an average MSE of 0.59% and an  $R^2$  of 98.86% for a three-day forecast, highlighting its resilience to varied parameter configurations. (3) In juxtaposition with prevalent models such as Support Vector Regression (SVR), K-Nearest Neighbors (KNN), Gradient Boosting Decision Tree (GBDT), XGBoost, Random Forest, and deep learning models like Deep Neural Network (DNN), Convolutional Neural Network (CNN), and standalone GRU-branch and Transformer-branch models, the GRU–Transformer framework demonstrates a significant advantage in predicting soil moisture content with enhanced precision for a five-day forecast. This underscores its exceptional capacity to navigate the intricacies of soil moisture data. This research not only provides a potent decision-support tool for agricultural irrigation planning but also makes a substantial contribution to the field of water resource conservation and optimization in agriculture, while concurrently imparting novel insights into the application of deep learning techniques in the spheres of agricultural and environmental sciences.

**Keywords:** GRU; transformer; soil moisture content; deep learning



**Citation:** Zheng, W.; Zheng, K.; Gao, L.; Zhangzhong, L.; Lan, R.; Xu, L.; Yu, J. GRU–Transformer: A Novel Hybrid Model for Predicting Soil Moisture Content in Root Zones. *Agronomy* **2024**, *14*, 432. <https://doi.org/10.3390/agronomy14030432>

Received: 13 January 2024

Revised: 18 February 2024

Accepted: 20 February 2024

Published: 23 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Maize, ubiquitously recognized as a cornerstone in global food agriculture and a prominent cash crop within the Chinese agrarian economy [1], holds a paramount position in the agricultural sector. In 2019, as per the Food and Agriculture Organization of the United Nations (FAO), China's maize yield escalated to an extraordinary 260 million tons, constituting a substantial 21.3% of the worldwide production [2]. However, the growth and fecundity of maize are contingent upon an array of determinants, with soil moisture content emerging as a critical environmental factor [3]. The hydration state of soil is pivotal, directly influencing the physiological and biochemical processes in maize, along with its root development [4]. Consequently, the precise prognostication of soil water content within the maize root zone is imperative for orchestrating efficient irrigation strategies and augmenting both the water-use efficiency and the crop's yield. Yet, forecasting the moisture content in maize's root zone presents a labyrinthine nonlinear challenge. It is intricately influenced by an amalgamation of factors such as meteorological conditions, soil characteristics, and various stages of crop growth, defying simplification into elementary mathematical models [5]. Thus, the exigent scientific quandary lies in harnessing the available meteorological and soil data to formulate a potent and efficacious predictive model.

Conventional methodologies for forecasting the soil moisture content in maize's root zone can be broadly categorized into three distinct classifications: physics-based approaches, statistics-based strategies, and machine learning-driven techniques [6]. Physics-based methods employ a combination of soil moisture movement and crop water demand equations, integrated with meteorological, soil, and crop parameters, to simulate the dynamic fluctuations of soil moisture. Exemplified by sophisticated models such as HYDRUS [7], SWAT (Soil and Water Assessment Tool) [8], DSSAT (Decision-support System for Agrotechnology Transfer) [9], and EPIC (Erosion Productivity Impact Calculator) [10], these methods adeptly mirror the physical underpinnings of soil moisture dynamics. However, they are hampered by their requirement for extensive parameter inputs, computational intricacy, and the need for highly precise parameter values, rendering them less viable for broad-scale and long-duration predictive applications [11]. Statistics-based methods, on the other hand, rely on the application of statistical principles and techniques to construct mathematical regression models. These models project future soil moisture levels based on historical meteorological and soil data [12] and are often represented through methodologies like linear regression [13], multiple regression [12], and exponential smoothing [14]. Despite their simplicity and ease of use, these methods fall short in acknowledging the nonlinear nature of soil moisture, failing to accurately track its dynamic variations and generally suffering from lower predictive accuracy [15]. Machine learning-based methods, meanwhile, harness the power of artificial intelligence to autonomously discern the intrinsic patterns of soil moisture from extensive meteorological and soil data, thereby predicting future soil moisture levels [13]. This category includes a diverse array of techniques such as the radial basis function neural network (RBF) [16], BP neural network [17], support vector machine (SVM) [18,19], extreme learning machine (ELM) [20], Random Forest [21], and Gradient Boosting Tree [22]. These methods are particularly proficient at processing nonlinear and high-dimensional data, yielding high prediction accuracy. However, they necessitate substantial volumes of training data, and their predictive performance is often limited by the models' inadequate feature extraction capabilities [23].

Deep learning, a sophisticated subset of machine learning, is founded on multi-layered neural networks capable of autonomously extracting features and discerning patterns from voluminous datasets, thereby boasting potent expressive and generalization proficiencies. The primary modalities within deep learning encompass the convolutional neural network (CNN) [24], the recurrent neural network (RNN) [25], and the self-attention mechanism [26]. As deep learning has evolved, methods such as DNN [27] and RNN [26] have been increasingly applied to the prediction of soil moisture content in agricultural contexts, achieving notable advancements. Nonetheless, traditional deep learning methodologies are not without their limitations. The CNN, utilizing convolutional kernels for feature extraction, is

adept at processing structured data like images and speech but falters with time-series data [28]. Conversely, the RNN, designed to handle time-series data, leverages historical information for future predictions but is plagued by gradient vanishing and explosion issues, impeding its ability to capture long-term dependencies [29]. Similarly, while the self-attention mechanism can effectively manage unstructured data such as natural language, it is less effective with structured data [30]. To address these challenges, the Gated Recurrent Unit (GRU) improves upon the RNN by incorporating a gating mechanism to mitigate gradient vanishing and explosion, thereby enhancing the model's memory capabilities [4]. Meanwhile, the Transformer, based on the self-attention mechanism, can discern global dependencies within an input sequence through multi-head attention and position encoding, significantly bolstering the model's expressive capacity [31]. Both GRU and Transformer, originating from the realm of natural language processing, excel in handling time-series prediction tasks and offer complementary advantages to redress the shortcomings of RNNs and self-attention mechanisms, respectively [32]. There have been successful applications of GRU and Transformer models in various fields, such as traffic flow analysis [33], speech recognition [34], and environmental sequence modeling [35], yet their application in predicting soil moisture content in maize's root zone remains relatively unexplored. Given the proven efficacy of hybrid models in enhancing prediction accuracy by amalgamating different deep learning approaches [4,36], this study proposes a novel technical solution: the creation of a hybrid model that fuses GRU and Transformer technologies. This GRU–Transformer hybrid model will be applied to the prediction of soil water content in the root zone of maize, with an aim to evaluate its feasibility and effectiveness in this specific domain.

The principal aim of this study is to introduce and assess the efficacy of a novel soil water content prediction model, christened the GRU–Transformer. This model's validation involves a rigorous evaluation using meteorological and soil moisture data collected from maize cultivation areas across eight distinct locations in Hebei Province, China, spanning the years 2011 to 2018. The assessment strategy encompasses an analysis of the model's predictive capabilities across various input configurations and forecast durations. Additionally, the performance of the GRU–Transformer model is benchmarked against a range of conventional machine learning and advanced deep learning models, providing a comprehensive understanding of its relative effectiveness in predicting soil moisture content in agricultural settings.

## 2. Material and Method

The hybrid modeling methodology delineated in this manuscript leverages the synergistic capabilities of GRU and Transformer algorithms to prognosticate future soil moisture levels across varying depths, as depicted in Figure 1. This innovative model is bifurcated into two distinct branches: the GRU branch and the Transformer branch. Each branch meticulously processes the 3D tensor, characterized by the dimensions (batch\_size, input\_length, num\_features), amalgamating input variables derived from both meteorological and soil datasets. Subsequently, the extracted features from each branch are intricately spliced and amalgamated through a meta-learner, culminating in the synthesis of the final predictive output. This output is represented as a two-dimensional tensor (batch\_size, output\_length), encapsulating the forecasted soil moisture content at a future juncture, thereby showcasing the model's robust predictive acumen.

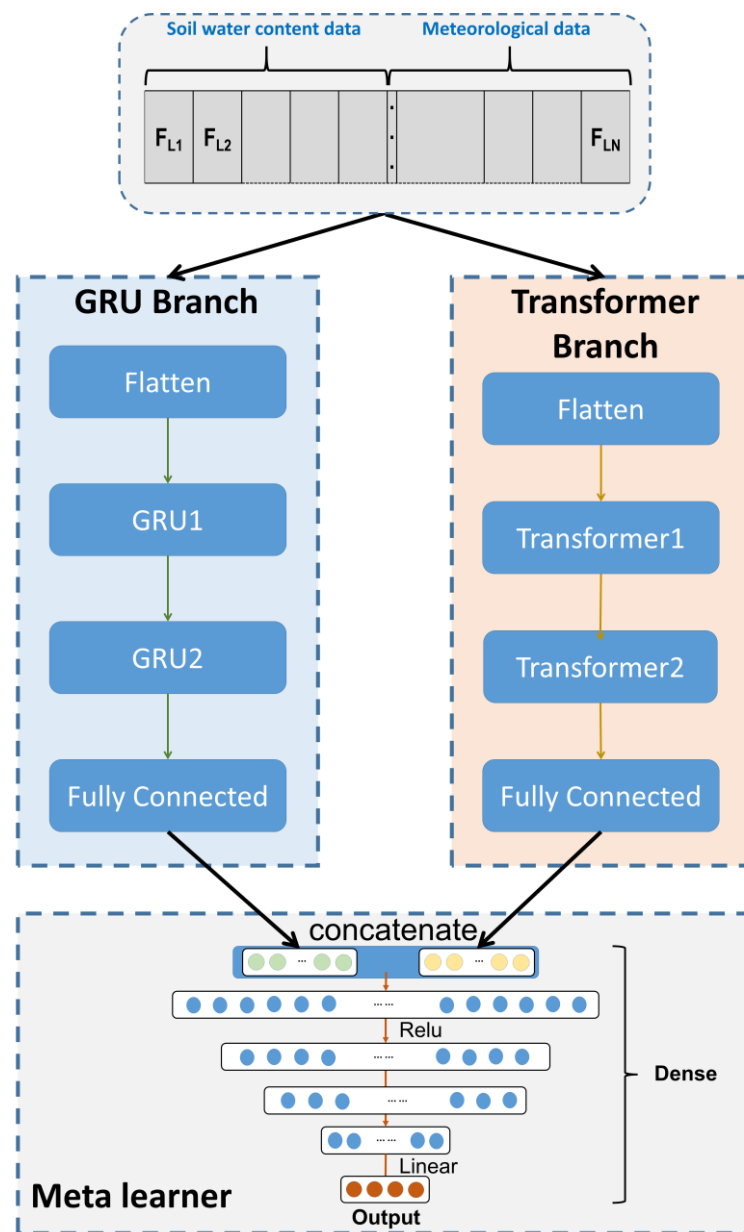


Figure 1. Architectural blueprint of the GRU–Transformer model.

### 2.1. GRU Branch

The purpose of the GRU branch in this study is to utilize the recurrent structure and gating mechanism of GRU to capture the long-term dependence and temporal dynamics features in the input sequences (Figure 2). GRU, using an improved RNN, can solve the problem of vanishing or exploding gradients that RNNs are prone to when dealing with long sequences [35]. The basic units of GRU are as follows:

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{1}$$

$$Z_t = \sigma(W_Z x_t + U_Z h_{t-1} + b_Z) \tag{2}$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \tag{3}$$

$$h_t = (1 - Z_t) \odot h_{t-1} + Z_t \odot \tilde{h}_t \tag{4}$$

where  $x_t$  is an input vector,  $h_t$  is an output vector,  $r_t$  is a reset gate,  $Z_t$  is an update gate,  $\sigma$  is a sigmoid function,  $\odot$  is an element-by-element multiplication,  $W_r, W_z, W_h, U_r, U_z, U_h$  is a learnable weight matrix, and  $b_r, b_z, b_h$  is a learnable bias vector. Reset gates and update gates can control the flow of information, selectively forgetting or remembering historical states. The reset gate can decide whether to reset the historical state or not, and the update gate can decide whether to update the current state or not.

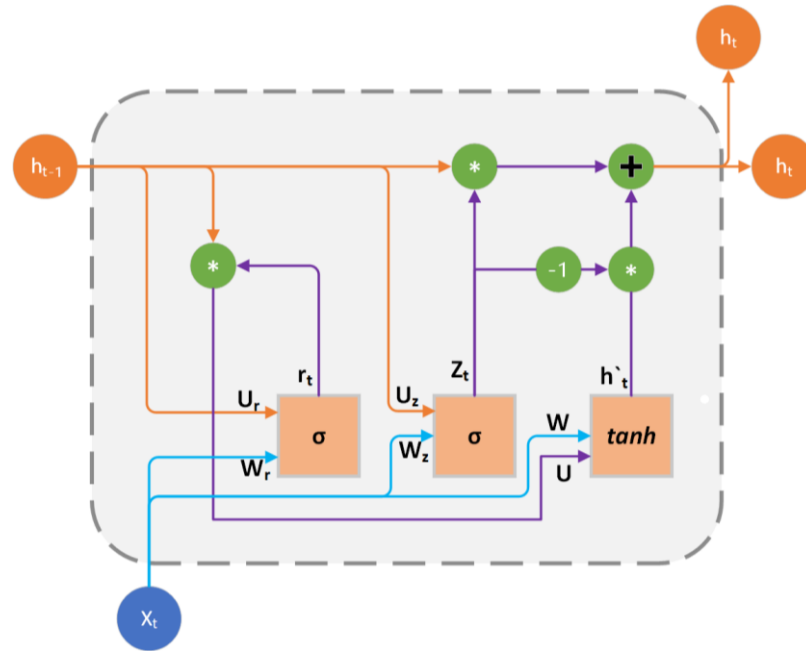


Figure 2. Conceptual illustration of the GRU branch.

The structural design and hyperparameter configuration of the GRU branch in the model proposed in this paper are shown in Table 1.

Table 1. Configuration of hyperparameters in the GRU branch.

Parameter	Value	Interpretation
Input dimension	22	Number of features received by the input layer
Network layer	2	Number of layers in the GRU network
Number of neurons per layer	512	Number of neurons per layer of GRU network
Activation function	ReLU	For increasing model depth and nonlinearity
Cyclic layer dropout ratio	0.2	Dropout ratios in the loop layer to prevent overfitting
Cyclic layer recurrent dropout ratio	0.2	Dropout ratios for internal connections in the loop layer
Number of neurons in the fully connected layer	256	The number of neurons to which the GRU output is mapped
Full connectivity layer dropout ratio	0.2	Dropout ratios for overfitting prevention in the fully connected layer
Fully connected layer activation function	Linear	Linear activation function for fully connected layers
Output layer dimension	256	Dimensions of GRU branch output

Firstly, a two-layer GRU network is constructed with 512 neurons in each layer. ReLU is used for the activation function to increase the depth and nonlinearity of the model and to improve the expressive power of the model. The formula for the ReLU function is

$$relu(x) = \max(0, x) \tag{5}$$

The output of the GRU is then mapped onto a low-dimensional space using a fully connected layer (Dense) of 256 neurons, reducing the number of parameters and decreasing the complexity of the model. The formula for the fully connected layer is

$$y = Wx + b \quad (6)$$

where  $x$  is the input vector,  $y$  is the output vector,  $W$  is the learnable weight matrix, and  $b$  is the learnable bias vector.

Finally, a spreading layer (Flatten) is connected to obtain a 2D tensor ( $batch\_size, 256$ ), which converts the output of the fully connected layer into a 1D vector for easy splicing with the output of the Transformer branch. The formula for the Flatten layer is

$$y = x.reshape(batchsize, -1) \quad (7)$$

where  $x$  is the input tensor,  $y$  is the output vector,  $batchsize$  is the batch size, and  $-1$  indicates that the remaining dimensions are computed automatically.

The output of the GRU branch can be expressed as

$$y = Flatten(Dense(GRU(GRU(x)))) \quad (8)$$

where  $x$  is the input tensor,  $y$  is the output vector,  $GRU$  is the GRU layer,  $Dense$  is the fully connected layer, and  $Flatten$  is the spreading layer.

## 2.2. Transformer Branch

In the proposed GRU–Transformer model, the Transformer branch plays a pivotal role in capturing global dependencies and multidimensional features within the input sequence. This is achieved through the employment of the Transformer’s self-attention and multi-head attention mechanisms, as depicted in Figure 3. The Transformer is adept at handling sequence-to-sequence tasks, utilizing its attention-based encoder–decoder structure [33]. This structure is integral to the model’s ability to process and analyze complex sequences of data effectively. The core unit of the Transformer can be conceptualized as follows:

$$Q = XW_Q \quad (9)$$

$$K = XW_K \quad (10)$$

$$V = XW_V \quad (11)$$

$$A = softmax\left(\frac{QK^T}{\sqrt{d_K}}\right) \quad (12)$$

$$Y = AV \quad (13)$$

where  $X$  is the input matrix,  $Q$ ,  $K$ ,  $V$  is the query matrix,  $W_Q$ ,  $W_K$ ,  $W_V$  key matrix, value matrix,  $A$  is the learnable weight matrix,  $Y$  is the attention matrix,  $d_K$  is the output matrix,  $\sqrt{d_K}$  is the dimension of the attention header,  $softmax$  is a scaling factor used to prevent the attention weights from being too large or too small, and  $softmax$  is the normalization function used to compute the weight of the value corresponding to each key. The attention mechanism calculates the correlation of each element in the input sequence with the other elements, thus enabling global dependency capture.

The structure of the Transformer branch in the model proposed in this paper is designed as follows: firstly, position information is added to each element in the input sequence by adding position embeddings to the inputs, so that the model can distinguish between elements at different positions. The detailed hyperparameter configuration is shown in Table 2.

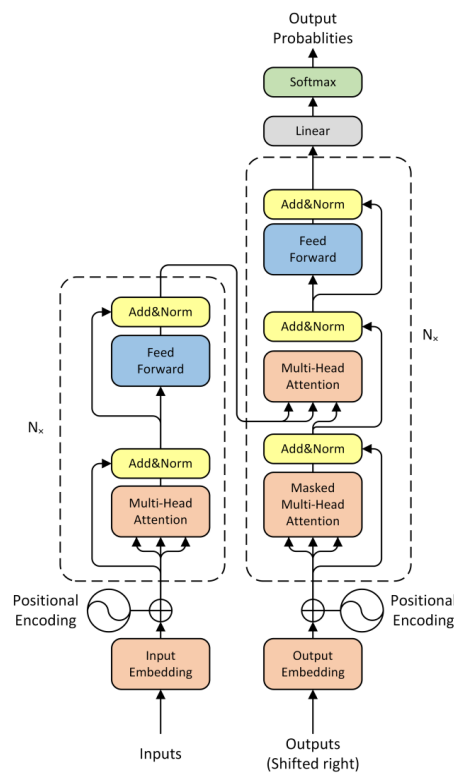


Figure 3. Conceptual diagram of the Transformer branch.

Table 2. Hyperparameter configuration for the Transformer branch.

Parameter	Value	Interpretation
Input dimension	22	Number of features received by the input layer
Number of encoder layers	4	Layers of the Transformer encoder
Attention headcount	4	Number of attention heads per layer of Transformer encoder
Dimensions of the attention head	64	Feature dimensions handled by each attention head
Dimensions of feedforward networks	128	Dimension of feedforward network in Transformer encoder
Dropout layer ratio	0.1	Ratio of Dropout layers used to prevent overfitting
Layer normalized epsilon	$1.00 \times 10^{-6}$	Small values in layer normalization to prevent division by zero
Number of neurons in the fully connected layer	256	Number of neurons to map the output of the Transformer
Full connectivity layer dropout ratio	0.2	Dropout ratios for overfitting prevention in the fully connected layer
Fully connected layer activation function	Linear	Linear activation function for fully connected layers
Output layer dimension	256	Dimension of Transformer branch output

The formula for position encoding is as follows:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \tag{14}$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \tag{15}$$

where  $PE$  is the position encoding matrix,  $pos$  is the position index,  $i$  is the dimension index,  $d$  is the dimension of the input. The dimensions of the position encoding are the same as the dimensions of the input and can be added directly to the input.

Then, four Transformer encoder layers (Transformer\_encoder) are constructed with four attention heads per layer (num\_heads = 4), each with a dimension of 64 (head\_size = 64), a feedforward network with a dimension of 128 (ff\_dim = 128), and a Dropout layer with a rate of 0.1 (dropout\_rate = 0.1). The above configuration increases the depth and diversity of the model and improves its expressive power. Among them, the Dropout layer can randomly discard a small number of neurons to prevent model overfitting.

The structure of the Transformer encoder layer is as follows:

$$X' = X + MultiHeadAttention(X, X, X) \tag{16}$$

$$X'' = X' + FeedForward(X') \tag{17}$$

$$Y = LayerNorm(X'') \tag{18}$$

where  $X$  is the input matrix,  $X'$  is the output of the first residual connection,  $X''$  is the output of the second residual connection,  $Y$  is the final output, *MultiHeadAttention* is the multi-head attention layer, *FeedForward* is the feedforward network layer, and *LayerNorm* is the layer normalization layer.

The multi-head attention layer can divide the input matrix into multiple sub-matrices, which are stitched together after separate attention computations to capture the multi-dimensional features in the input sequence. The feedforward network layer can increase the complexity of the model by performing a nonlinear transformation of the input matrix. The layer normalization layer normalizes each layer of the input matrix to make the model more stable. Residual connections make the model easier to optimize avoiding gradient vanishing or exploding. Then, a 256 neuron fully connected layer (Dense) is constructed to map the output of the Transformer to a low dimensional space reducing the number of parameters and decreasing the complexity of the model. Finally, a two-dimensional tensor (batch\_size, 256) is obtained through a Flatten layer, which converts the output of the fully connected layer into a one-dimensional vector, which can be easily spliced with the output of the GRU branch.

The output of the Transformer branch can be represented as follows:

$$y = Flatten\left(Dense\left(TransformerEncoder^4(PositionEmbeddings(x))\right)\right) \tag{19}$$

where  $x$  is the input tensor,  $y$  is the output vector, *PositionEmbeddings* is the position encoding layer, *TransformerEncoder* is the Transformer encoder layer, *Dense* is the fully-connected layer, *Flatten* is the spreading layer, and 4 means repeat four times.

### 2.3. Meta-Learner

The purpose of the meta-learner in the model proposed in this paper is to splice and fuse the outputs of the two branches to obtain the final prediction results. The structure of the meta-learner is designed as follows: firstly, the outputs of the two branches are concatenated to obtain a two-dimensional tensor (batch\_size, 512), and then the features of the two branches are combined to increase the information content of the model. The detailed hyper-parameter configuration is shown in Table 3.

**Table 3.** Meta-learner hyperparameter configuration.

Parameter	Value	Interpretation
Dimensions after splicing	512	GRU and Transformer branches output spliced dimensions
Number of neurons in layers 1–5 (decreasing)	512, 256, 128, 64, 32	Number of neurons in the first five fully connected layers, decreasing layer by layer
Number of neurons in layer 6 (output layer)	6	Number of neurons in the 6th fully connected layer, corresponding to the output dimension
Dropout ratio per fully connected layer	0.2	Dropout ratios for overfitting prevention in each fully connected layer
Fully connected layer activation Function per layer	ReLU	Activation functions for fully connected layers except the last one
Output layer activation function	Linear	Linear activation function for the output layer

The equation for the splicing layer can be expressed as follows:

$$y = [x_1, x_2] \tag{20}$$

where  $x_1$  is the output of the GRU branch,  $x_2$  is the output of the Transformer branch,  $y$  is the spliced output, and  $[]$  denotes the splicing operation.



Then, a 5-layer neuron fully connected layer (Dense) is constructed and the activation function is ReLU. The formula for the fully connected layer is as follows:

$$y = \text{relu}(Wx + b) \quad (21)$$

where  $x$  is the input vector,  $y$  is the output vector,  $W$  is the learnable weight matrix,  $b$  is the learnable bias vector, and  $\text{relu}$  is the activation function.

Finally, the connection is made to the output layer, which is Dense, and the activation function used is linear to obtain the final output, which is a two-dimensional tensor (batch\_size, output\_length). The above process maps the output of the fully connected layer to a one-dimensional space to obtain the predicted values.

#### 2.4. Model Evaluation and Training

To evaluate the performance of different models, the following five evaluation metrics were chosen in this paper.

Mean Squared Error (MSE):

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (22)$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (23)$$

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (24)$$

Coefficient of Determination ( $R^2$ ):

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \quad (25)$$

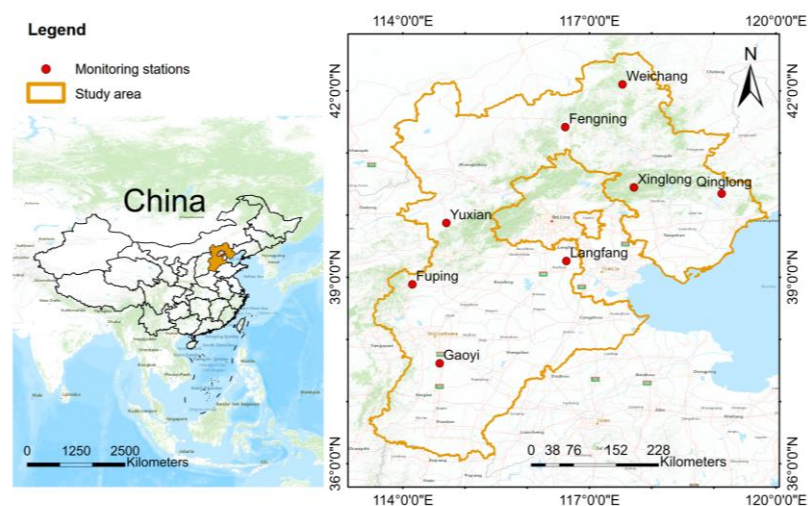
where  $\hat{y}_i$  is the predicted value,  $y_i$  is the true value,  $\bar{y}_i$  is the average of the true values. *MAE* can reflect the actual situation of the error of the predicted value: the smaller its value, the higher the accuracy of the prediction. *MSE* is the expected value of the square of the difference between the predicted value and the true value, which can evaluate the degree of change of the data, the smaller its value, the higher the stability of the prediction. *RMSE* is the arithmetic square root of *MSE*, which, like the *MAE*, is used to measure the accuracy of the prediction, but it has a greater penalty for larger errors, so it is more suitable for evaluating models that are sensitive to errors.  $R^2$  can eliminate the influence of dimensionality on the evaluation index, it indicates the degree of correlation between the predicted value and the real value, and the closer its value is to 1, the higher the fit of the prediction.

Regarding the experimental setup for this research, the hardware configuration encompasses an Intel® Xeon® CPU E5-1620 v4 @ 3.50 GHz, an NVIDIA Quadro K2200 GPU, and 32 GB of RAM. The software infrastructure leverages Anaconda as the foundational platform for deep learning endeavors, with Keras serving as the framework for constructing deep learning models, and TensorFlow-gpu 1.13 functioning as the backend engine. CUDA technology, developed by NVIDIA Corporation based in Santa Clara, CA, USA, is employed to facilitate parallel computing on the GPU, thereby augmenting the training speed of the model. Python 3.7 is utilized as the primary programming language.

The training methodology incorporates the Adam optimization algorithm for model refinement, complemented by an early stopping criterion as a termination condition for training. This criterion is set with a threshold of 50 iterations—training ceases if there is no improvement in the model’s loss on the validation set within these iterations. The optimal model weights are preserved, and the model is saved in the .h5 format. This comprehensive setup ensures a robust and efficient training environment, which is crucial for the development and validation of the deep learning model.

### 2.5. Study Area and Data Acquisition

The focal area of this study is Hebei Province, China, encompassing geographical coordinates ranging from  $36^{\circ}05'$  to  $42^{\circ}40'$  N latitude and  $113^{\circ}27'$  to  $119^{\circ}50'$  E longitude. This region is characterized by a temperate continental monsoon climate, distinguished by four well-defined seasons, ample sunshine, moderate rainfall, and significant temperature variations, among other climatic attributes. The dataset employed in this study was procured from the China Meteorological Data Network, encompassing both meteorological and soil moisture content data spanning the years 2011 to 2018. In alignment with the prevalent conditions of maize cultivation in the region and subsequent consultations with local agricultural authorities, a comprehensive dataset comprising 450,120 records was assembled from eight agrometeorological monitoring stations within Hebei Province (namely Fuping, Weixian, Gaoyi, Fengning, Langfang, Xinglong, Weichang, and Qinglong stations), as illustrated in Figure 4. Soil water content measurements were conducted at an hourly frequency. The volumetric water content of the soil (denoted as 10SVWC, 20SVWC, 30SVWC, 40SVWC, 50SVWC, and 60SVWC, represented in percentage) was recorded at varying depths of 10 cm, 20 cm, 30 cm, 40 cm, 50 cm, and 60 cm, with daily average values being computed. The meteorological data encompassed a range of parameters, including the daily mean air temperature (TEM\_Avg, in degrees Celsius), minimum air temperature (TEM\_Min, in degrees Celsius), maximum air temperature (TEM\_Max, in degrees Celsius), mean surface temperature (GST\_Avg, in degrees Celsius), minimum surface temperature (GST\_Min, in degrees Celsius), maximum surface temperature (GST\_Max, in degrees Celsius), duration of sunshine (SSH, in hours), mean relative humidity (RHU\_Avg, in percentages), precipitation during the period from 20:00 to 08:00 h (PRE\_Time\_2008, in millimeters), precipitation during the period from 08:00 to 20:00 h (PRE\_Time\_0820, in millimeters), 24 h precipitation (PRE\_Time\_2020, in millimeters), average wind speed (WIN\_S\_Avg, in kilometers per hour), maximum wind speed (WIN\_S\_Max, in kilometers per hour), dominant wind direction (WIN\_D\_S\_Max, in degrees), instantaneous maximum wind speed (WIN\_S\_Inst\_Max, in kilometers per hour), and the direction of the instantaneous maximum wind speed (WIN\_D\_INST\_Max, in degrees).



**Figure 4.** Geographical overview of the study area and distribution of sites.

## 2.6. Data Analysis

Table 4 elegantly presents the synthesized statistical analysis of 22 distinct soil and meteorological indicators, encompassing an array of parameters such as mean, standard deviation, minimum, maximum, and median values across the eight meticulously chosen monitoring stations. The temporal scope of this study encompasses an entire annual cycle of data from maize cultivation locales, thereby encapsulating the complete spectrum of seasonal fluctuations and chronological diversities. This comprehensive temporal coverage, while providing a holistic view of the data, simultaneously introduces a formidable challenge in the precise prediction of soil water content at varying depths within the maize root zone. Such an encompassing approach ensures a thorough understanding of the intricate interplay between temporal variables and the dynamic nature of soil moisture content, critical for advancing predictive accuracy in agricultural environmental studies.

**Table 4.** Comprehensive statistical analysis outcomes of soil and meteorological data.

Indicator	Abbreviation	Unit	Mean	Standard Deviation	Minimum Value	Maximum Value	Median
10 cm Soil Volumetric Water Content	10SVWC	%	13.2	6.64	2.2	53.2	12.1
20 cm Soil Volumetric Water Content	20SVWC	%	14.21	7.4	2.3	65.9	14
30 cm Soil Volumetric Water Content	30SVWC	%	15.09	6.92	3.1	62	15
40 cm Soil Volumetric Water Content	40SVWC	%	16.4	7.53	3.5	64	16.7
50 cm Soil Volumetric Water Content	50SVWC	%	15	7.27	3.6	52.5	14
60 cm Soil Volumetric Water Content	60SVWC	%	17.68	8.21	3.8	61.4	17
Average temperatures	TEM_Avg	°C	11.79	11.19	−24.0	32.7	13.7
Maximum temperature	TEM_Max	°C	18.39	11.26	−21.4	40.9	20.8
Minimum temperature	TEM_Min	°C	6.24	11.29	−27.8	28.6	7.5
Average relative humidity	RHU_Avg	%	58.46	19.64	8	100	59
Periods of precipitation in 20-08	PRE_Time_2008	mm	9.64	56.78	0	2495	0
8–20 h precipitation period	PRE_Time_0820	mm	9.26	48.81	0	1678	0
20-20 precipitation period	PRE_Time_2020	mm	22.53	347.03	0	2495	0
Average wind speed	WIN_S_Avg	km/h	20.21	10.53	2	123	18
Maximum wind speed	WIN_S_Max	km/h	53.14	22.85	12	213	49
Maximum wind direction	WIN_D_S_Max	°	9.41	4.63	1	16	10
Maximum wind speed	WIN_S_Inst_Max	km/h	88.93	37.95	19	393	82
Maximum wind direction	WIN_D_INST_Max	km/h	9.36	4.66	1	16	10
Daylight hours	SSH	h	6.96	3.94	0	14.3	7.9
Average ground temperature	GST_Avg	°C	14.59	12.84	−21.9	41.1	16.6
Maximum ground temperature	GST_Max	°C	32.25	16.71	−14.1	71.6	33.9
minimum ground temperature	GST_Min	°C	4.81	11.71	−27.2	27.2	5.6

Figure 5 elegantly delineates the outcomes of the Pearson correlation analysis conducted among the various variables. This analysis reveals that the correlation coefficients pertaining to soil water content at diverse depths surpass the threshold of 0.7, indicating a robust and strongly positive correlation. The interplay between each meteorological parameter and soil water content exhibits varying degrees of correlation. Specifically, TEM\_Avg, TEM\_Max, TEM\_Min, RHU\_Avg, PRE\_Time\_2008, PRE\_Time\_0820, PRE\_Time\_2020, GST\_Avg, GST\_Max, and GST\_Min all demonstrate a positive correlation with the soil water content at different depths. Conversely, WIN\_S\_Avg, WIN\_S\_Max, WIN\_D\_S\_Max, WIN\_S\_Inst\_Max, WIN\_D\_INST\_Max, and SSH exhibit a negative correlation with soil water content across various depths. This intricate mosaic of correlations underscores the multifaceted nature of the interactions between meteorological factors and soil water content, providing invaluable insights into the complex dynamics governing soil moisture across different strata.

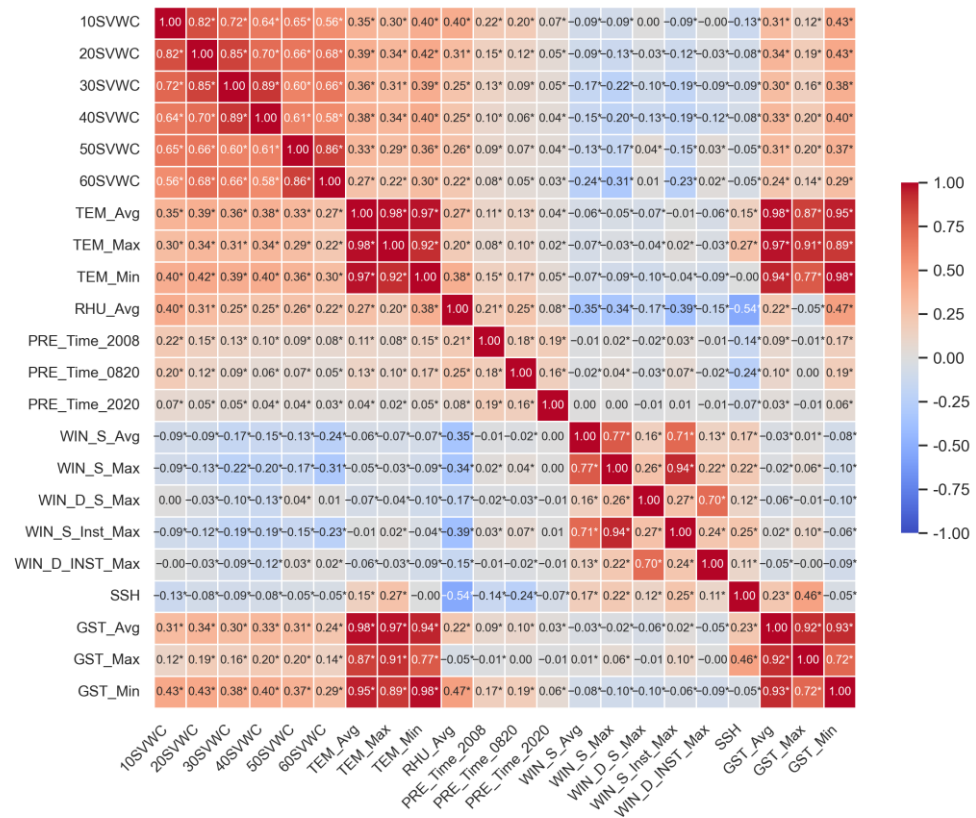


Figure 5. Analytical depiction of correlation between soil water content and other variables at varied depths. Asterisks (\*) indicate correlations statistically significant at  $p < 0.05$ .

### 3. Results and Discussion

#### 3.1. Comparison of Model Prediction Accuracy for Different Days of Delay

Figures 6 and 7 elucidate the predictive accuracy outcomes of the GRU–Transformer model across various delay intervals. The model’s overall performance is noteworthy for its consistently high accuracy and stability under diverse conditions. Specifically, the MSE for soil moisture prediction at different depths fluctuated between 2.71% and 17.53%, while the  $R^2$  spanned from 60.47% to 96.08%. These figures emphatically underscore the model’s adaptability and resilience in varied scenarios. A more granular analysis reveals that the model’s predictive precision is notably superior for shorter delay periods, such as 1 to 2 days. For instance, the average MSE for a 1-day delay stands at 5.22%, with the lowest recorded value being 2.71%, and the average  $R^2$  is remarkably high at 89.92%. Nonetheless, it is observed that the predictive error incrementally ascends with an increase in the delay duration, suggesting a heightened proficiency of the model in forecasting near-term data. When delving into the comparison across different soil depths, it is discernible that the model’s predictive error for shallow soils, like at 10 cm depth, is relatively elevated, with a mean MSE of 15.00% and an average  $R^2$  of merely 66.17%. In stark contrast, the model’s error rate for deeper soils, such as at 50 cm, is markedly lower, boasting an average MSE of 7.71% and an average  $R^2$  of 85.66%. This disparity highlights the model’s significant advantage in predicting the water content of deeper soils. In summation, the soil water content prediction model introduced in this study exhibits exemplary efficacy across various prediction delays and soil depths. It is particularly commendable for its high accuracy and reliability in short-term predictions and in assessing the water content of deeper soil layers.

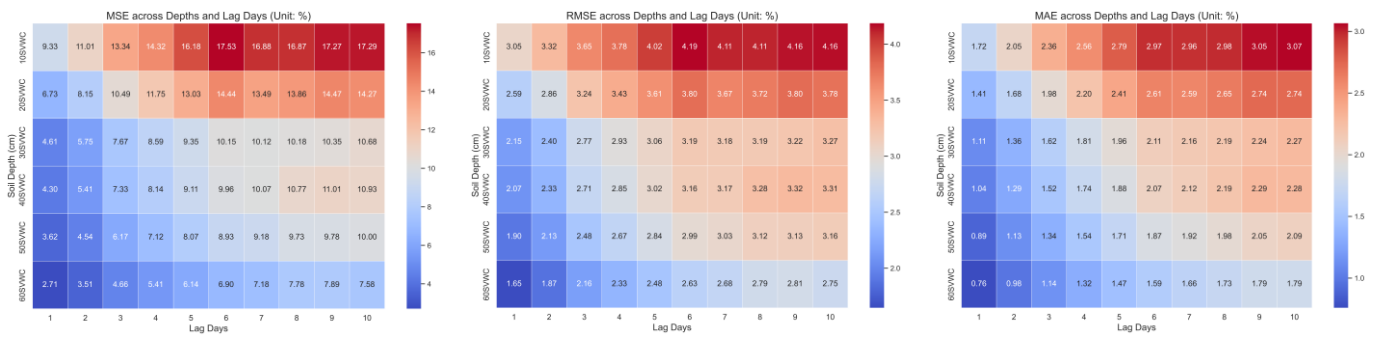


Figure 6. Analysis of prediction error in GRU–Transformer model across different time delays.

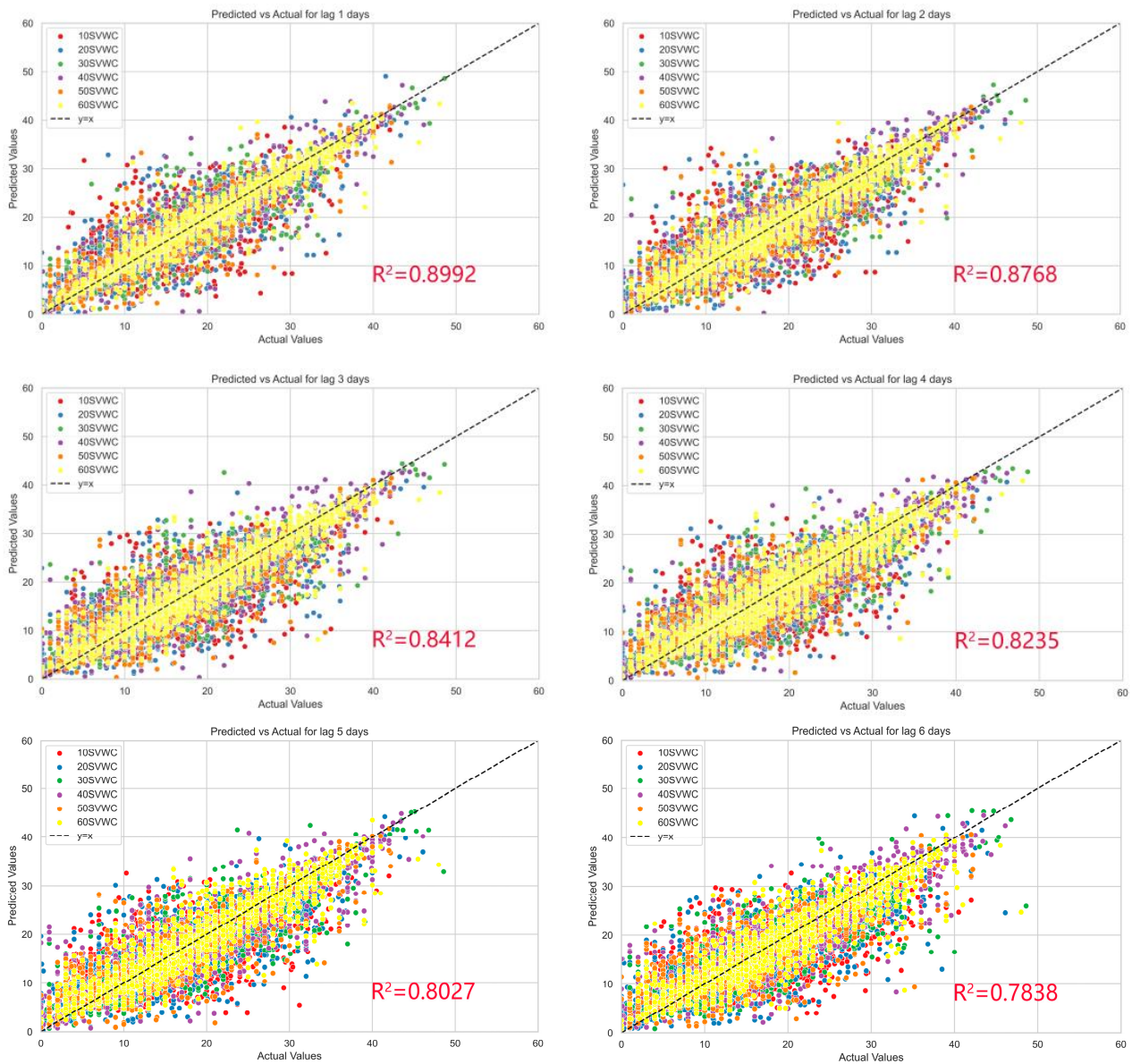
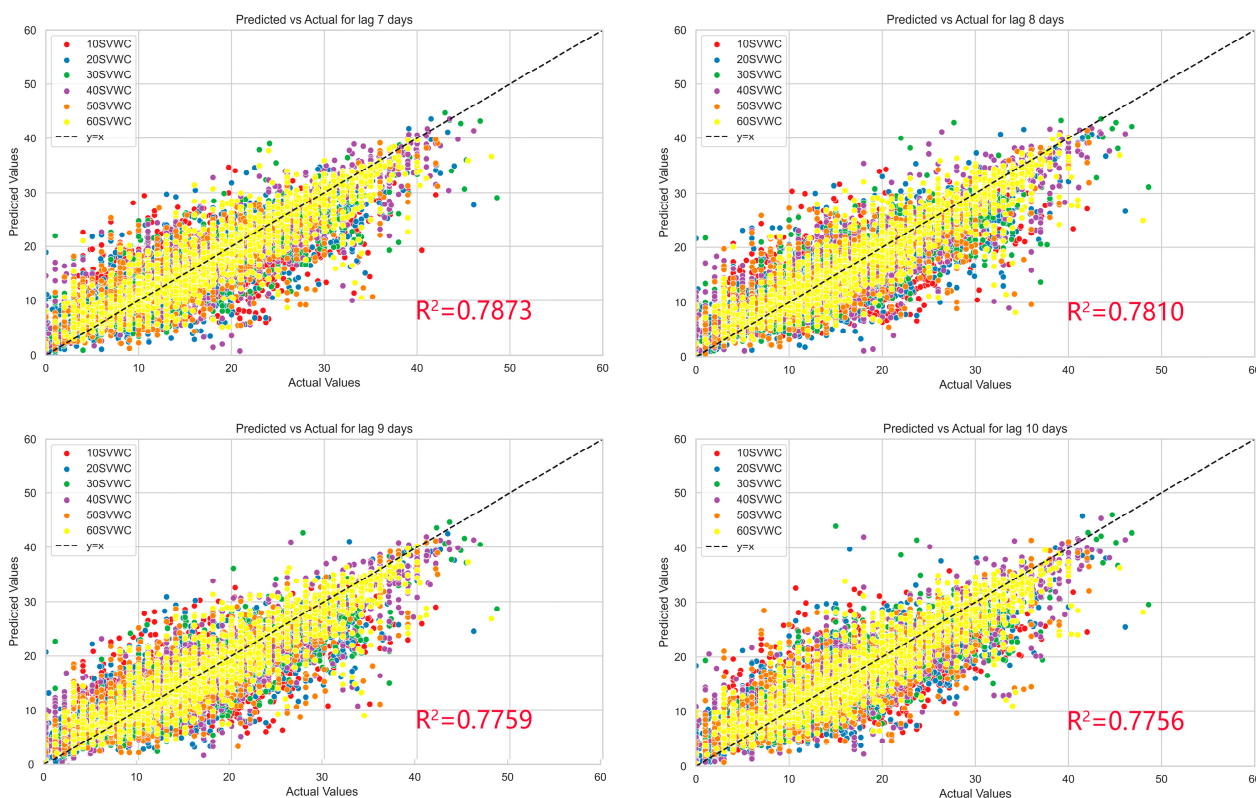


Figure 7. Cont.

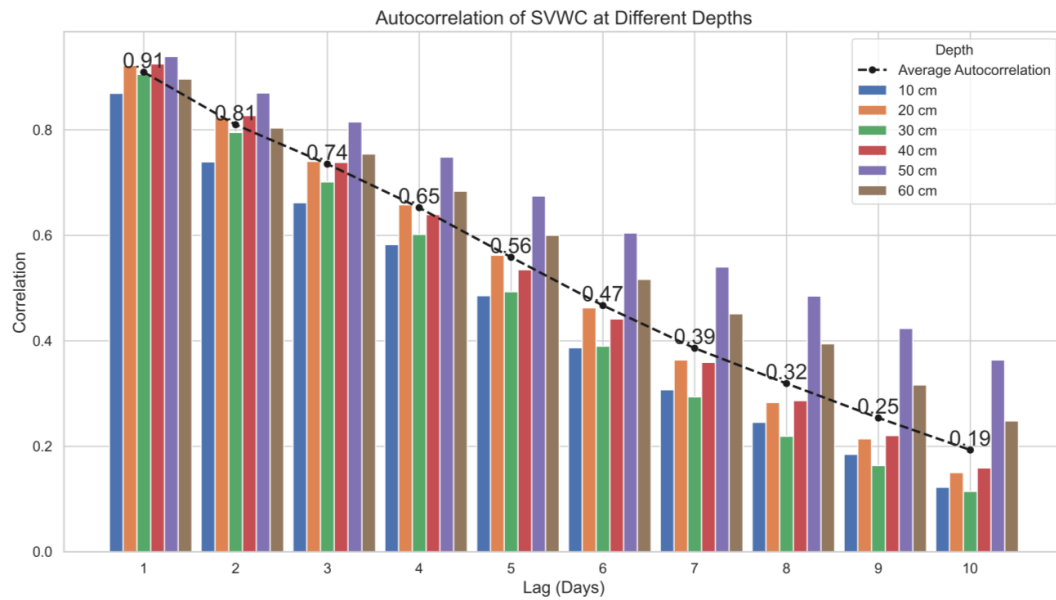


**Figure 7.** Error distribution in test set for GRU–Transformer model over various time delays.

The marked disparities in the performance of the soil moisture content prediction model at varying delay intervals and soil depths, as highlighted in this study, underscore the intricate complexity of soil moisture dynamics and its susceptibility to environmental influences. These variances not only attest to the model’s adaptability but also furnish novel insights into the understanding of soil moisture behavior. The observed decrease in prediction accuracy with prolonged delay days is potentially attributable to escalating uncertainties on the temporal scale. In the immediate term, the soil moisture content is relatively less impacted by meteorological conditions, thereby enabling the model to forecast near-future soil moisture status with greater precision. However, the predictability diminishes as the timeframe extends, largely due to the increasing unpredictability of meteorological factors such as rainfall and evaporation, which exert a direct influence on the model’s accuracy [37]. Figure 8 delineates the outcomes of an autocorrelation analysis on soil water content over a 1–10-day delay period, using the data from this study. The analysis reveals a decline in the mean autocorrelation coefficient of soil water content from a peak of 0.91 to a mere 0.19 as the delay interval increases, thereby highlighting the amplified challenges in forecasting over longer future durations. This phenomenon aligns with the findings of Cai et al. (2019) [29], who examined the 1–16-day autocorrelation of soil water content.

Furthermore, the dynamics of soil moisture encompass a spectrum of intricate non-linear processes such as evaporation, infiltration, and plant water uptake. Consequently, the model’s prediction error tends to escalate with the extension of the forecast period, a concept corroborated by theoretical studies in time-series prediction. From the standpoint of soil depth, the model exhibits heightened accuracy in predicting moisture content in deeper soils. This heightened accuracy could stem from the reduced impact of daily climatic fluctuations on deeper soils, which exhibit relatively less variability in moisture content. In contrast, surface soils are directly exposed to atmospheric elements, rendering their moisture status more immediately susceptible to variables like rainfall and evapotranspiration, thus posing greater challenges for accurate prediction [38]. The moisture dynamics in

deeper soil strata are comparatively stable and are more effectively and accurately captured by existing hydrological and soil science models [39]. This observation underscores the significant role of soil physical characteristics in influencing the performance of prediction models and suggests the necessity of comprehensive consideration of the properties of different soil layers in the construction and parameter optimization of these models.



**Figure 8.** Autocorrelation analysis of soil water content under 1–10-day delay.

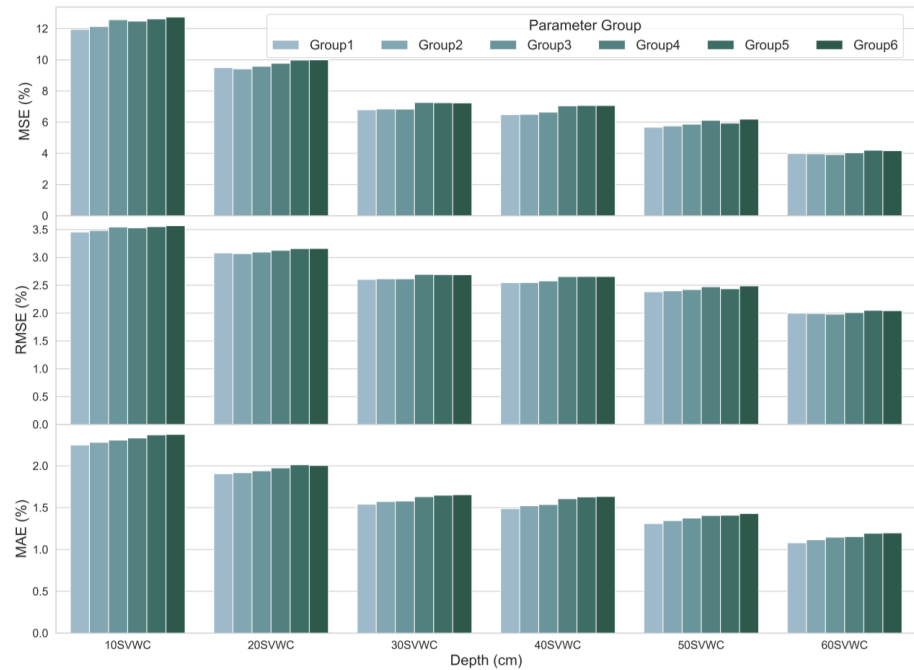
### 3.2. Assessing the Predictive Efficacy of Varied Input Parameter Combinations

The efficacy of the proposed soil water content prediction model was exhaustively evaluated across a spectrum of input parameter combinations. Six distinct sets of input terms, comprising 6, 9, 12, 15, 18, and 22 parameters, respectively, were meticulously formulated based on the inter-correlation of the indicators (as outlined in Table 5). This approach was employed to juxtapose the model’s predictive accuracies under diverse parameter amalgamations. The experimental outcomes elucidate key characteristics and trends in the model’s performance.

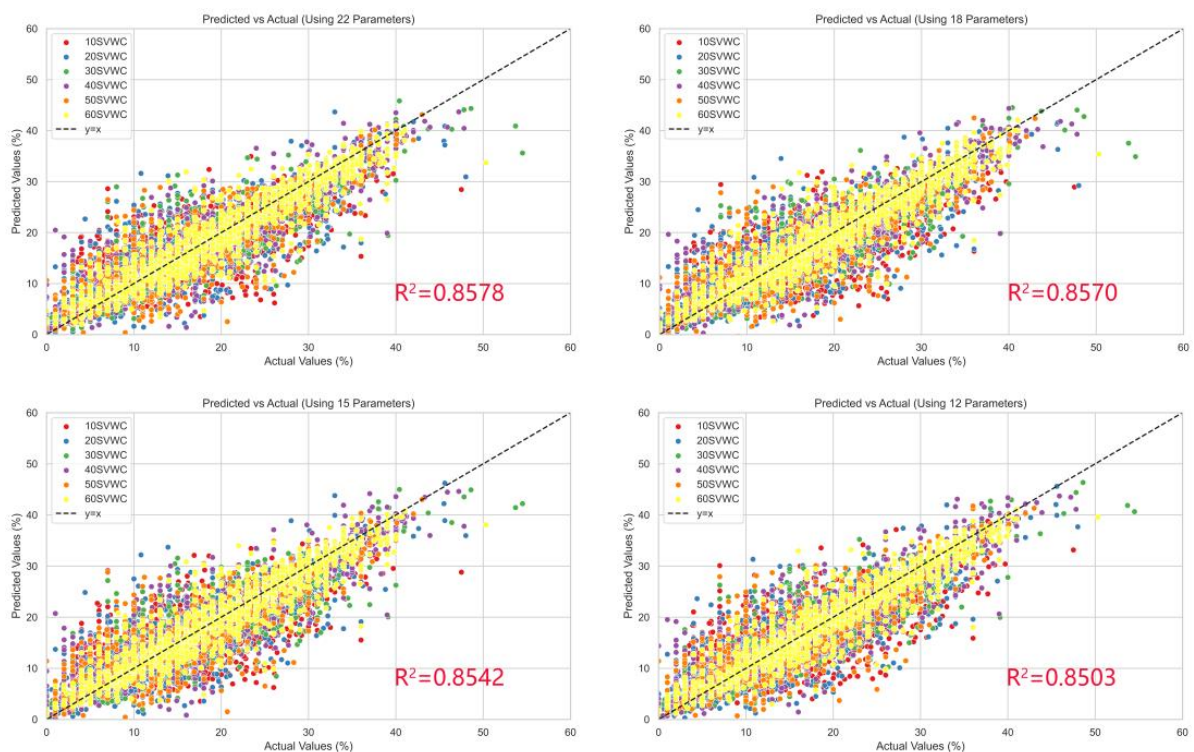
**Table 5.** Framework of diverse input term combinations.

Combination	Number of Parameters	Parameters
Group 1	22	10SVWC, 20SVWC, 30SVWC, 40SVWC, 50SVWC, 60SVWC, TEM_Avg, TEM_Max, TEM_Min, RHU_Avg, PRE_Time_2008, PRE_Time_0820, PRE_Time_2020, WIN_S_Avg, WIN_S_Max, WIN_D_S_Max, WIN_S_Inst_Max, WIN_D_INST_Max, SSH, GST_Avg, GST_Max, GST_Min
Group 2	18	10SVWC, 20SVWC, 30SVWC, 40SVWC, 50SVWC, 60SVWC, TEM_Avg, TEM_Max, TEM_Min, RHU_Avg, PRE_Time_2008, PRE_Time_0820, PRE_Time_2020, WIN_S_Avg, WIN_S_Max, WIN_D_S_Max, WIN_S_Inst_Max, WIN_D_INST_Max
Group 3	15	10SVWC, 20SVWC, 30SVWC, 40SVWC, 50SVWC, 60SVWC, TEM_Avg, TEM_Max, TEM_Min, RHU_Avg, PRE_Time_2008, PRE_Time_0820, PRE_Time_2020, WIN_S_Avg, WIN_S_Max
Group 4	12	10SVWC, 20SVWC, 30SVWC, 40SVWC, 50SVWC, 60SVWC, TEM_Avg, TEM_Max, TEM_Min, RHU_Avg, PRE_Time_2008, PRE_Time_0820
Group 5	9	10SVWC, 20SVWC, 30SVWC, 40SVWC, 50SVWC, 60SVWC, TEM_Avg, TEM_Max, TEM_Min
Group 6	6	10SVWC, 20SVWC, 30SVWC, 40SVWC, 50SVWC, 60SVWC

Figures 9 and 10 depict the predictive accuracy of the GRU–Transformer model under these varied input scenarios. Regarding its overall performance, the model consistently exhibited high accuracy and stability across the range of parameter combinations. The MSE of the GRU–Transformer model oscillated between 0.43% and 1.11%, while the  $R^2$  ranged from 0.8482 to 0.8578. This variance indicates that the proposed model adeptly predicts soil water content, effectively handling different sets of input parameters.

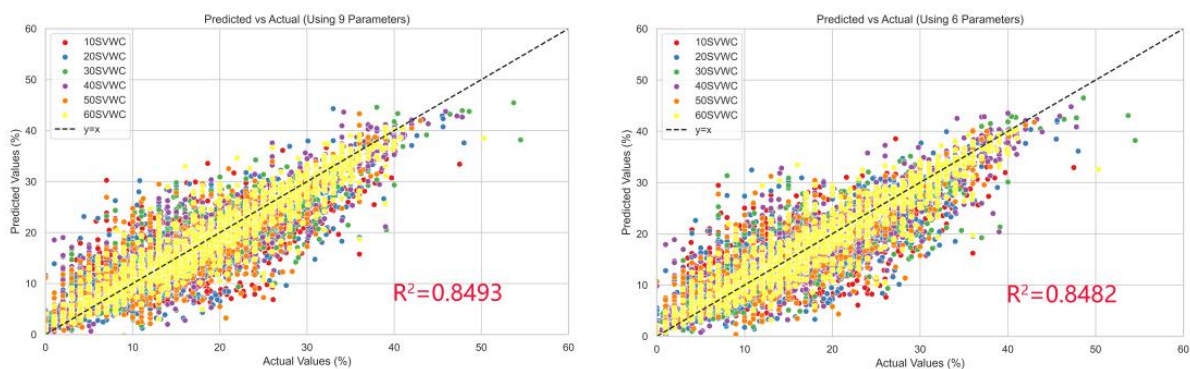


**Figure 9.** Assessment of prediction accuracy in GRU–Transformer model with varied input term combinations.



**Figure 10.** Cont.





**Figure 10.** Distribution of test set errors in GRU–Transformer model under different input term combinations.

Focusing on the comparative performance across parameter combinations, the model attained its zenith under the 22-parameter ensemble (Group 1), registering an average *MSE* of 0.62% and an average  $R^2$  of 0.8578. This implies that the model harnesses a broader array of data features to enhance predictive accuracy when endowed with a higher count of parameters. Conversely, in the 6-parameter combination (Group 6)—the minimal parameter set—the average *MSE* modestly declined to 0.59%, and the average  $R^2$  marginally ascended to 0.8482. This demonstrates the model’s capacity to sustain high accuracy even with a reduced parameter set, underlining its robustness and adaptability.

An in-depth analysis of the influence of the parameter count on the model’s efficacy reveals that, while both the maximal and minimal *MSE* values diminish as the number of parameters decreases, the overall performance remains relatively unaffected. This observation suggests that the model proficiently captures the salient features of soil water content, even under conditions of simplified input, highlighting its capability to efficiently distill and utilize essential data characteristics.

This research discerned that diminishing the number of parameters led to a slight escalation in the error indices, while the  $R^2$  values continued to hover at elevated levels. This phenomenon suggests that the model retains the primary factors influencing soil water content, even under a regime of parameter simplification, thereby facilitating effective prediction. Such a capability could be attributed to the fact that the chosen parameters exhibit a strong correlation with soil water content, coupled with a degree of redundancy among them. That is, certain parameters can be inferred or approximated using others. Consequently, when the parameter count is reduced, the model compensates for the absence of certain parameters by adjusting the weights of the remaining ones, thus sustaining its high predictive accuracy. This characteristic also underscores the model’s robust self-adaptive nature, enabling automatic adjustments in its structure and parameters in response to varying parameter combinations, to achieve optimal prediction outcomes.

Moreover, the progressive increase in *MSE*, *RMSE*, and *MAE* from Group 1 (22 parameters) to Group 6 (6 parameters), juxtaposed with the relatively minor fluctuation in the  $R^2$  value (0.8482–0.8578), indicates the model’s tolerance to changes in parameter quantity. This implies that alterations in the number of parameters do not significantly impact the predictive capability of the model. This resilience may stem from the model’s GRU–Transformer hybrid neural network architecture, which adeptly extracts higher-order features of the parameters through multilayer nonlinear transformations, thereby enhancing its expressive and generalization abilities. Hence, even with a limited number of parameters, the model is capable of discerning potential interrelationships through neural network learning, facilitating accurate soil water content prediction.

The high consistency in error metrics between Groups 1 and 2 may be attributed to the inclusion of comprehensive soil volumetric water content parameters and other critical parameters such as temperature, relative humidity, precipitation, and wind speed in these combinations. These parameters collectively encapsulate the soil’s physical, chemical, and

biological attributes, as well as the soil–atmosphere interactions, thereby influencing the dynamic shifts in soil water content. Hence, they provide a wealth of information that enhances the predictive accuracy and stability of the model.

Interestingly, even with Group 6, which comprises the least number of parameters focused solely on volumetric soil water content, the decline in predictive accuracy is not significantly pronounced. This implies that these parameters are direct indicators of soil water content and the primary factors influencing its variability. Additionally, as depicted in Figure 11, the correlation between different soil depths ranges from 0.65 to 0.95, offering synergistic features that bolster the stability of the model’s predictions. Consequently, the model achieves satisfactory predictive accuracies even when solely reliant on soil water content data.

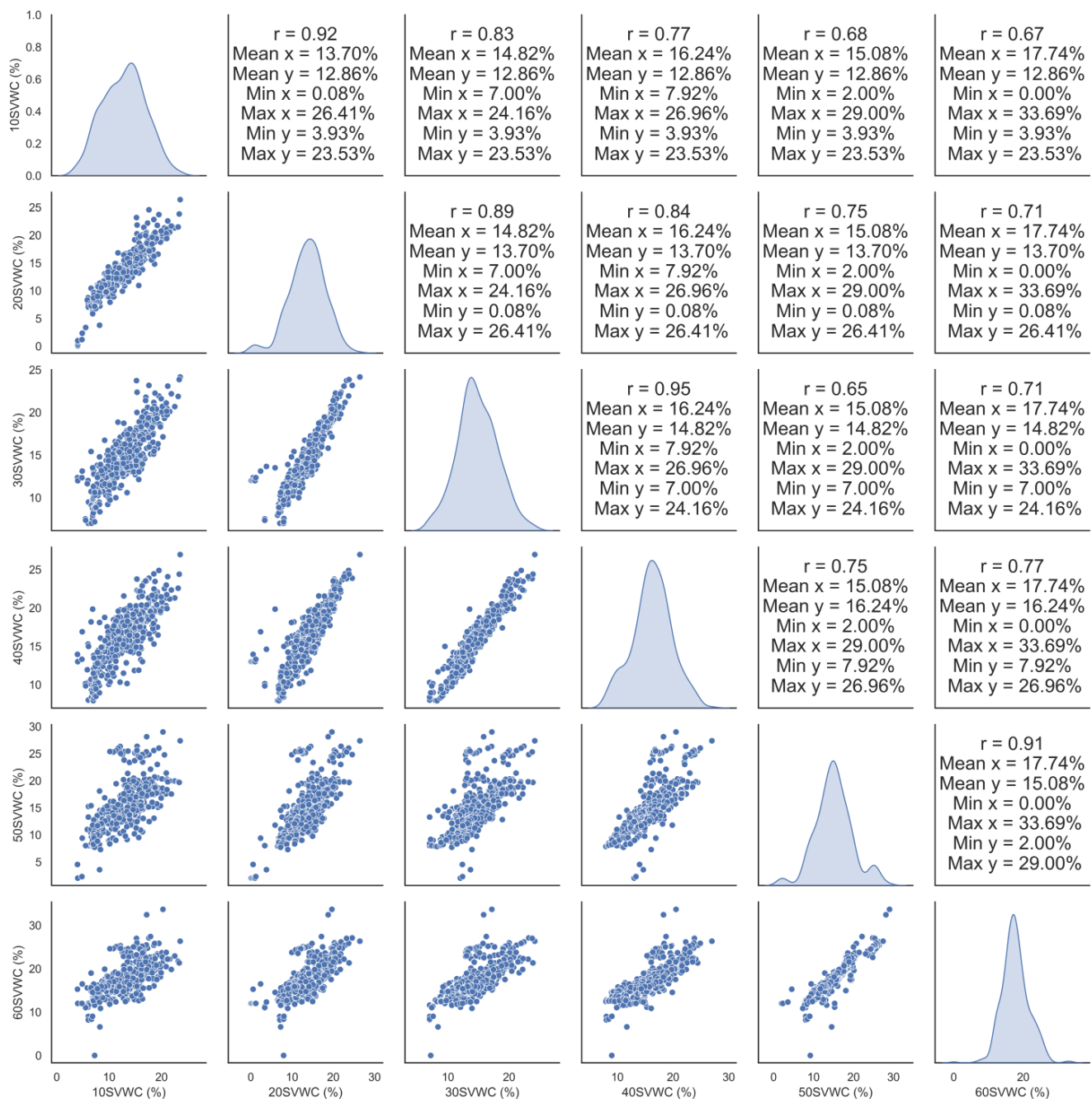


Figure 11. Characterization and interrelation of soil water content at diverse depths.

### 3.3. Comparative Analysis of Prediction Accuracy against Benchmark Models

In assessing the performance for five-day-ahead predictions, this study scrutinizes the distinctions in the predictive accuracy of the newly devised GRU–Transformer model relative to other established models. This comparative evaluation encompasses typical machine learning models such as Support Vector Regression (SVR [18]), K-Nearest Neighbors (KNN [40]), Gradient Boosted Decision Trees (GBDT [41]), XGBoost [42], Random Forest [21]), and deep learning models like Deep Neural Networks (DNN [27]), Convolutional Neural Networks (CNN [4]), as well as independent GRU and Transformer branches.

Table 6 delineates the soil water content prediction accuracies of these various models, showcasing a spectrum of performances with *MSE* values ranging from 10.31% to 31.40% and  $R^2$  values spanning 42.32% to 80.27%. These results highlight the diverse capabilities and adaptability of the different models in predicting soil water content.

**Table 6.** Comparative analysis of prediction accuracy across various models for soil water content.

Model	<i>MSE</i> (%)	<i>RMSE</i> (%)	<i>MAE</i> (%)	$R^2$
KNN	31.40	5.60	4.26	0.4232
SVR	22.58	4.74	3.55	0.5844
GBDT	19.59	4.39	2.56	0.6274
DNN	18.78	4.2	3.17	0.6454
XGBoost	12.69	3.52	2.28	0.7579
LSTM	12.52	3.51	2.36	0.7627
Random Forest	12.20	3.46	2.35	0.7676
CNN	11.94	3.42	2.27	0.7725
GRU branch	11.86	3.41	2.21	0.7738
Transformer branch	11.04	3.29	2.18	0.7895
GRU–Transformer	10.31	3.17	2.04	0.8027

Upon comparing the performances of these models, the GRU–Transformer model, as proposed in this study, emerged as the most proficient, recording an *MSE* of 10.31% and an  $R^2$  of 80.27%. This exemplifies its superior predictive capabilities in handling complex soil moisture data. In stark contrast, the traditional KNN model fared the poorest, with an *MSE* of 31.40% and an  $R^2$  of merely 42.32%. The XGBoost model, known for its efficacy and popularity, yielded an *MSE* of 12.69% and an  $R^2$  of 75.79%, outperforming most models but still falling short of the GRU–Transformer model’s proficiency.

When comparing deep learning models against traditional machine learning models, it is observed that deep learning-based models (such as CNN, LSTM, independent GRU, and Transformer branches) generally surpass their traditional counterparts. Their *MSE* and  $R^2$  values lie in the range of 11% to 13% and 76% to 79%, respectively, underscoring the formidable capability of deep learning in processing complex datasets. Conversely, traditional machine learning models, including SVR and GBDT, exhibit comparatively lower predictive accuracy, with significantly higher *MSE* and *MAE* values than those of deep learning models. This dichotomy accentuates the burgeoning potential of deep learning approaches in the realm of complex data analysis and prediction.

In the realm of soil water content prediction, the exemplary performance of the GRU–Transformer model underscores the formidable potential of deep learning technologies in processing environmental data. The model’s capacity to surpass both traditional machine learning and other deep learning models can be primarily attributed to its unique structural composition and its enhanced capability in handling complex datasets. The GRU–Transformer model ingeniously amalgamates two cutting-edge neural network architectures: the GRU and the Transformer. This combination allows for an effective capture of long-term dependencies and intricate spatial features within time-series data. The GRU component of the model is particularly adept at processing time-series data [4], proving especially potent in comprehending and predicting the dynamically evolving nature of soil water content. Meanwhile, the Transformer branch, through its self-attention mecha-

nism [35], excels in deciphering complex interrelationships between different data points, thereby substantially elevating the model's prediction accuracy.

In contrast to conventional machine learning models like SVR and GBDT, deep learning models demonstrate a heightened ability to manage highly nonlinear and multi-dimensional data. This study reveals that deep learning models, such as CNNs and LSTM networks, generally outperform traditional machine learning approaches, highlighting deep learning's prowess in capturing intricate data patterns and relationships [27]. However, the GRU-Transformer model elevates this capability to a higher echelon by synergizing the strengths of both GRU and Transformer architectures. By integrating these two powerful neural network structures, their individual advantages are mutually amplified [33].

This synergistic fusion not only allows the GRU-Transformer model to surpass the accuracy of individual deep learning models but also to outshine traditional machine learning methods. This enhancement is pivotal in the context of environmental data processing, where the complexity and variability of data demand sophisticated analytical approaches. The GRU-Transformer model, with its dual-architecture advantage, thus stands as a significant advancement in the field, pushing the boundaries of what can be achieved in soil water content prediction and environmental data analysis.

### *3.4. Limitations and Further Study*

We advocate the utilization of deep learning methodologies for the intricate processing of meteorological and soil data, with the aim of achieving precise predictions of soil moisture at various depths. This approach effectively harnesses the strengths of the two models to enhance predictive outcomes. Nevertheless, this study is not devoid of limitations, which present opportunities for refinement in future endeavors. A primary shortcoming is the relatively limited dataset size, encompassing merely eight consecutive years of data from eight locations. This constraint poses a potential risk of overfitting, potentially impairing the model's generalization capabilities. Additionally, this study did not incorporate factors such as the growth stages of maize or other agricultural management practices, including irrigation and fertilization, all of which can significantly influence soil moisture dynamics.

In light of these limitations, future research directions will focus on expansion and optimization in several key areas. (1) Enlarging the dataset: Future studies will aim to broaden the dataset's scope, encompassing data from diverse regions, encompassing different years and maize varieties. This expansion will provide a more comprehensive and varied dataset, enhancing the robustness and applicability of the model. (2) Incorporating additional variables: The inclusion of variables such as the maize's growth stage, various agricultural management practices, and soil types will be a focal point. These additions will offer a more holistic view of the factors impacting soil moisture, thus enriching the model's predictive accuracy. (3) Model optimization: We plan to refine the structure and parameters of the model further. This will involve exploring various combination methods and fusion strategies, with an emphasis on enhancing the model's efficiency and stability. Through these enhancements, we aim to not only address the current study's limitations but also significantly advance the field of soil moisture prediction using deep learning techniques, ultimately contributing to more effective and informed agricultural management. Given its demonstrated accuracy and versatility, the GRU-Transformer model holds potential for application in a range of environmental contexts, including the prediction of flash droughts, thereby contributing valuable insights for proactive disaster management and environmental preservation.

## **4. Conclusions**

This research introduces a novel hybrid modeling approach that synergizes GRU and Transformer architectures, aiming to predict soil moisture content at varying depths, specifically within the maize root zone. Utilizing meteorological and soil moisture data from eight maize cultivation sites in Hebei Province, China, spanning from 2011 to 2018, the study meticulously evaluates the model's predictive performance across diverse input

combinations and varying forecast durations. This evaluation also includes comparative analysis with conventional machine learning and other deep learning models. The findings of this study reveal distinct prediction efficacies under varying conditions. The results showed that (1) the GRU–Transformer model exhibited significant advantages under different prediction delay days and soil depth conditions. The model demonstrated the highest accuracy in the short-term prediction of 1 to 2 days, in which the mean *MSE* was 5.22% and the lowest reached 2.71% in the prediction with a 1-day delay, while the mean  $R^2$  was as high as 89.92%. The prediction accuracy decreased as the number of days of delay in prediction increased, but the overall performance remained stable. In addition, in the comparison of different soil depths, the model showed higher accuracy in predicting the water content of deeper soil layers, which was much better than the performance of shallow soil prediction. The methodology and experimental design adopted in the study fully considered the adaptability and robustness of the model under different environmental conditions. (2) By comparing the model performance under different combinations of input parameters, we found that the prediction error of the model rises with the decrease in parameters; however, even under the combination of only soil water content parameters, the model predicts the third day in the future with an average *MSE* of 0.59% and an average  $R^2$  of 98.86%, which reflects the high tolerance of the model to different combinations of parameters. (3) In comparative analysis, the GRU–Transformer model substantially outperformed a range of typical machine learning models (such as SVR, KNN, GBDT, XGBoost, and Random Forest) and deep learning models (including DNN, CNN, independent GRU, and Transformer branches). This superiority affirms the model's exceptional capability in handling complex soil moisture data. Conclusively, this study presents a potent decision-support tool for agricultural irrigation management, contributing significantly to enhancing the efficiency and conservation of agricultural water resources. The GRU–Transformer model, with its robust predictive accuracy and adaptability, offers valuable insights for the sustainable development of agriculture, addressing critical needs in environmental data processing and agricultural resource management.

**Author Contributions:** Methodology, W.Z., J.Y. and K.Z.; investigation, K.Z. and L.G.; validation, K.Z. and J.Y.; formal analysis, J.Y.; resources, L.X.; data curation, K.Z. and L.G.; writing—original draft preparation, K.Z.; writing—review and editing, J.Y.; visualization, K.Z.; supervision, L.X., R.L. and L.Z.; project administration, J.Y.; funding acquisition, L.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project was supported by the National Key Research and Development Program (2022YFD1900803), the Yunnan Provincial Basic Research Plan Project (202101AT070248), the Beijing Academy of Agriculture and Forestry Sciences Youth Research Fund (QNJJ202410), and the Earmarked Fund for CARS-02 and CARS-54.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Rigden, A.; Mueller, N.; Holbrook, N.; Pillai, N.; Huybers, P. Combined Influence of Soil Moisture and Atmospheric Evaporative Demand Is Important for Accurately Predicting US Maize Yields. *Nat. Food* **2020**, *1*, 127–133. [CrossRef] [PubMed]
2. China's Corn Production Increases in 2021 | World Grain. Available online: <https://www.world-grain.com/articles/16188-chinas-corn-production-increases-in-2021> (accessed on 13 January 2024).
3. Li, L.; Li, X.; Zheng, X.; Li, X.; Jiang, T.; Ju, H.; Wan, X. The Effects of Declining Soil Moisture Levels on Suitable Maize Cultivation Areas in Northeast China. *J. Hydrol.* **2022**, *608*, 127636. [CrossRef]
4. Yu, J.; Xin, Z.; Xu, L.; Dong, J.; Zhangzhong, L. A Hybrid CNN-GRU Model for Predicting Soil Moisture in Maize Root Zone. *Agric. Water Manag.* **2020**, *245*, 106649. [CrossRef]
5. Babaeian, E.; Sadeghi, M.; Jones, S.; Montzka, C.; Vereecken, H.; Tuller, M. Ground, Proximal and Satellite Remote Sensing of Soil Moisture. *Rev. Geophys.* **2019**, *57*, 530–616. [CrossRef]

6. Simunek, J.; Van Genuchten, M.T.; Šejna, M. HYDRUS: Model Use, Calibration, and Validation. *Trans. ASAE Am. Soc. Agric. Eng.* **2012**, *55*, 1261–1274. [[CrossRef](#)]
7. Er-Raki, S.; Ezzahar, J.; Merlin, O.; Amazirh, A.; Ait Hssaine, B.; Kharrou, H.; Khabba, S.; Chehbouni, A. Performance of the HYDRUS-1D Model for Water Balance Components Assessment of Irrigated Winter Wheat under Different Water Managements in Semi-Arid Region of Morocco. *Agric. Water Manag.* **2020**, *244*, 106546. [[CrossRef](#)]
8. Belmans, C.; Wesseling, J.; Feddes, R. *Simulation Model of the Water Balance of a Cropped Soil Providing Different Types of Boundary Conditions (SWATRE)*; ICW: Wageningen, The Netherlands, 1981; Volume 63.
9. Awan, Z.; Khaliq, T.; Akhtar, M.; Imran, A.; Irfan, M.; Ahmed, M.J.; Ahmad, A. Building Climate-Resilient Cotton Production System for Changing Climate Scenarios Using the DSSAT Model. *Sustain. Sci.* **2021**, *131*, 10495. [[CrossRef](#)]
10. Wang, Z.; Ye, L.; Jiang, J.; Fan, Y.; Zhang, X. Review of Application of EPIC Crop Growth Model. *Ecol. Model.* **2022**, *467*, 109952. [[CrossRef](#)]
11. Zheng, W.; Zhangzhong, L.; Xin, Z.; Wang, C.; Sun, S.; Niu, H. A Review on the Soil Moisture Prediction Model and Its Application in the Information System. In Proceedings of the 11th IFIP WG 5.14 International Conference, CCTA 2017, Jilin, China, 12–15 August 2017; Springer: Cham, Switzerland, 2019; pp. 352–364, ISBN 978-3-030-06136-4.
12. Aguilera, H.; Moreno-Merino, L.; Wesseling, J.; Jiménez-Hernández, M.; Castaño Castaño, S. Soil Moisture Prediction to Support Management in Semiarid Wetlands during Drying Episodes. *Catena* **2016**, *147*, 709–724. [[CrossRef](#)]
13. Acharya, U.; Daigh, A.; Oduor, P. Machine Learning for Predicting Field Soil Moisture Using Soil, Crop, and Nearby Weather Station Data in the Red River Valley of the North. *Soil Syst.* **2021**, *5*, 57. [[CrossRef](#)]
14. Fang, H.; Zhang, Y.; Wei, S.; Li, W.; Ye, Y.; Sun, T.; Weiwei, L. Validation of Global Moderate Resolution Leaf Area Index (LAI) Products over Croplands in Northeastern China. *Remote Sens. Environ.* **2019**, *233*, 111377. [[CrossRef](#)]
15. Ahansal, Y.; Bouziani, M.; Yaagoubi, R.; Sebari, I.; Sebari, K.; Kenny, L. Towards Smart Irrigation: A Literature Review on the Use of Geospatial Technologies and Machine Learning in the Management of Water Resources in Arboriculture. *Agronomy* **2022**, *12*, 297. [[CrossRef](#)]
16. Chen, C.; Tan, J.; Yin, J.; Zhang, F.; Yao, J. Prediction for Soil Moisture in Tobacco Fields Based on PCA and RBF Neural Network. *Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agric. Eng.* **2010**, *26*, 85–90. [[CrossRef](#)]
17. Yang, X.; Zhang, C.; Cheng, Q.; Zhang, H.; Gong, W. A Hybrid Model for Soil Moisture Prediction by Using Artificial Neural Networks. *Rev. De La Fac. De Ing.* **2017**, *32*, 265–271.
18. Fan, J.; Yue, W.; Wu, L.; Zhang, F.; Cai, H.; Xiukang, W.; Lu, X.; Xiang, Y. Evaluation of SVM, ELM and Four Tree-Based Ensemble Models for Predicting Daily Reference Evapotranspiration Using Limited Meteorological Data in Different Climates of China. *Agric. For. Meteorol.* **2018**, *263*, 225–241. [[CrossRef](#)]
19. Hong, Z.; Kalbarczyk, Z.; Iyer, R. A Data-Driven Approach to Soil Moisture Collection and Prediction. In Proceedings of the 2016 IEEE International Conference on Smart Computing (SMARTCOMP), St. Louis, MO, USA, 1 May 2016; pp. 1–6.
20. Prasad, R.; Deo, R.; Li, Y.; Maraseni, T. Soil Moisture Forecasting by a Hybrid Machine Learning Technique: ELM Integrated with Ensemble Empirical Mode Decomposition. *Geoderma* **2018**, *330*, 136–161. [[CrossRef](#)]
21. Carranza, C.; Nolet, C.; Peziz, M.; Ploeg, M. Root Zone Soil Moisture Estimation with Random Forest. *J. Hydrol.* **2020**, *593*, 125840. [[CrossRef](#)]
22. Wu, L.; Fan, J. Comparison of Neuron-Based, Kernel-Based, Tree-Based and Curve-Based Machine Learning Models for Predicting Daily Reference Evapotranspiration. *PLoS ONE* **2019**, *14*, e0217520. [[CrossRef](#)] [[PubMed](#)]
23. Chen, W.; Zheng, Z.; Yu, J.; Wang, C.; Huang, R. Data-Driven Calibration of Soil Moisture Sensor Considering Impacts of Temperature: A Case Study on FDR Sensors. *Sensors* **2019**, *19*, 4381. [[CrossRef](#)] [[PubMed](#)]
24. Kim, T.-Y.; Cho, S.-B. Predicting Residential Energy Consumption Using CNN-LSTM Neural Networks. *Energy* **2019**, *182*, 72–81. [[CrossRef](#)]
25. Khaki, S.; Wang, L.; Archontoulis, S. A CNN-RNN Framework for Crop Yield Prediction. *Front. Plant Sci.* **2020**, *10*, 1750. [[CrossRef](#)] [[PubMed](#)]
26. Li, Q.; Zhu, Y.; Shangguan, W.; Wang, X.; Li, L.; Yu, F. An Attention-Aware LSTM Model for Soil Moisture and Soil Temperature Prediction. *Geoderma* **2022**, *409*, 115651. [[CrossRef](#)]
27. Cai, Y.; Zheng, W.; Xin, Z.; Zhangzhong, L.; Xue, X. Research on Soil Moisture Prediction Model Based on Deep Learning. *PLoS ONE* **2019**, *14*, e0214508. [[CrossRef](#)]
28. Sajedian, I.; Kim, J.; Rho, J. Finding the Optical Properties of Plasmonic Structures by Image Processing Using a Combination of Convolutional Neural Networks and Recurrent Neural Networks. *Microsyst. Nanoeng.* **2019**, *5*, 27. [[CrossRef](#)]
29. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [[CrossRef](#)]
30. Chen, Y.; Bruzzone, L.; Jiang, L.; Sun, Q. ARU-Net: Reduction of Atmospheric Phase Screen in SAR Interferometry Using Attention-Based Deep Residual U-Net. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5780–5793. [[CrossRef](#)]
31. Zhong, Z.; Li, Y.; Ma, L.; Li, J.; Zheng, W.-S. Spectral-Spatial Transformer Network for Hyperspectral Image Classification: A Factorized Architecture Search Framework. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–15. [[CrossRef](#)]
32. Shewalkar, A.N. *Comparison of RNN, LSTM and GRU on Speech Recognition Data*; North Dakota State University: Fargo, ND, USA, 2018.

33. Zhang, Y.; Liu, S.; Zhang, P.; Li, B. GRU- and Transformer-Based Periodicity Fusion Network for Traffic Forecasting. *Electronics* **2023**, *12*, 4988. [[CrossRef](#)]
34. Aloysius, N.; Madathilkulangara, G.; Nedungadi, P. Incorporating Relative Position Information in Transformer-Based Sign Language Recognition and Translation. *IEEE Access* **2021**, *9*, 145929–145942. [[CrossRef](#)]
35. Chen, D.; Yongchareon, S.; Lai, E.; Yu, J.; Sheng, Q.; Li, Y. Transformer with Bidirectional GRU for Nonintrusive, Sensor-Based Activity Recognition in a Multiresident Environment. *IEEE Internet Things J.* **2022**, *9*, 23716–23727. [[CrossRef](#)]
36. Yu, J.; Tang, S.; Zhangzhong, L.; Zheng, W.; Wang, L.; Wong, A.; Xu, L. A Deep Learning Approach for Multi-Depth Soil Water Content Prediction in Summer Maize Growth Period. *IEEE Access* **2020**, *8*, 199097–199110. [[CrossRef](#)]
37. Zhai, C.; Zhou, H.; Zhao, J. Experimental Study on Inter-Annual Water Requirement and Water Consumption of Drip Irrigation Maize in North of Xinjiang. *Sci. Agric. Sin.* **2017**, *50*, 2769–2780. [[CrossRef](#)]
38. Guo, J.-L.; Yin, G.-H.; Gu, J.; Liu, Z.-X. Determination of Irrigation Scheduling of Spring Maize in Different Hydrological Years in Fuxin, Liaoning Province Based on CROPWAT Model. *Chin. J. Ecol.* **2016**, *35*, 3428–3434. [[CrossRef](#)]
39. Zhou, S.; Hu, X.; Wang, W.; Zhang, Y. Water-Saving and Stable Yield Effects of Regulation on Soil Wetted Depth in Different Growth Stage of Spring Maize. *Trans. Chin. Soc. Agric. Eng.* **2016**, *32*, 125–132.
40. Ghawi, R.; Pfeffer, J. Efficient Hyperparameter Tuning with Grid Search for Text Categorization Using kNN Approach with BM25 Similarity. *Open Comput. Sci.* **2019**, *9*, 160–180. [[CrossRef](#)]
41. Ahmadianfar, I.; Bozorg-Haddad, O.; Chu, X. Gradient-Based Optimizer: A New Metaheuristic Optimization Algorithm. *Inf. Sci.* **2020**, *540*, 131–159. [[CrossRef](#)]
42. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13 August 2016; pp. 785–794.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.