

Article

Machine Learning-Based Estimation of Tractor Performance in Tillage Operations Using Soil Physical Properties

So-Yun Gong ¹, Seung-Min Baek ² , Seung-Yun Baek ^{2,3} , Yong-Joo Kim ^{2,4,*}  and Wan-Soo Kim ^{1,5,6,*} 

¹ Department of Bio-Industrial Machinery Engineering, College of Agriculture and Life Sciences, Kyungpook National University, Daegu 41566, Republic of Korea; sunny722@knu.ac.kr

² Eco-Friendly Hydrogen Electric Tractor & Agricultural Machinery Institute, Chungnam National University, Daejeon 34134, Republic of Korea; bsm1104@naver.com (S.-M.B.); kelpie0037@gmail.com (S.-Y.B.)

³ Department of Agricultural and Biosystems Engineering, North Dakota State University, Fargo, ND 58102, USA

⁴ Department of Smart Agriculture Systems Machinery Engineering, Chungnam National University, Daejeon 34134, Republic of Korea

⁵ Department of Smart Bio-Industrial Mechanical Engineering, Kyungpook National University, Daegu 41566, Republic of Korea

⁶ Upland Field Machinery Research Center, Kyungpook National University, Daegu 41566, Republic of Korea

* Correspondence: babina@cnu.ac.kr (Y.-J.K.); wansoo.kim@knu.ac.kr (W.-S.K.); Tel.: +82-42-821-6716 (Y.-J.K.); +82-53-950-5794 (W.-S.K.)

Abstract

Accurate estimation of tractor performance under various soil conditions is essential for enhancing operational efficiency in precision agriculture. This study developed machine learning models to estimate tractor performance based on key soil physical properties. Three algorithms—decision tree (DT), CatBoost, and LightGBM—were employed to capture nonlinear relationships between soil parameters and tractor performance indicators. The input variables included soil moisture content, cone index, and particle composition, while the output variables were engine torque, power, slip ratio, and axle power. The models in this study were trained and validated using field data collected from eight paddy fields in Chungcheongnam-do (two in Seosan, two in Cheongyang, and four in Dangjin) and two paddy fields in Gyeonggi-do (Anseong), Republic of Korea. Results showed that models using multiple soil variables significantly outperformed those using single variables. In Model D, CatBoost demonstrated superior performance in predicting engine torque, engine power, slip ratio, and axle power, achieving R^2 values that were 7.0–14.2% higher than those of DT and 1.6–3.8% higher than those of LightGBM. These findings demonstrate the feasibility of using machine learning with minimal input data to estimate tractor performance, potentially reducing the reliance on extensive physical testing.



Academic Editors: Anna Rita Bernadette Cammerino and Massimo Monteleone

Received: 28 August 2025

Revised: 11 September 2025

Accepted: 19 September 2025

Published: 21 September 2025

Citation: Gong, S.-Y.; Baek, S.-M.; Baek, S.-Y.; Kim, Y.-J.; Kim, W.-S.

Machine Learning-Based Estimation of Tractor Performance in Tillage Operations Using Soil Physical Properties. *Agronomy* **2025**, *15*, 2228. <https://doi.org/10.3390/agronomy15092228>

Copyright: © 2025 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/>).

Keywords: agricultural tractor; load measurement system; machine learning; soil physical properties; tillage performance

1. Introduction

Tractors are indispensable agricultural machines employed in nearly all stages of farming operations, including plowing, harrowing, seeding, and harvesting [1,2]. Consequently, accurately evaluating their operational performance and energy efficiency is essential. Key performance indicators—such as engine torque (ET), engine power (EP), slip ratio (SR), and axle power (AP)—serve as vital metrics for assessing driving performance, fuel efficiency, agricultural implement suitability, and the overall economic viability of

farming operations [3,4]. Prior research has consistently underscored the importance and practical applicability of these performance indicators [5].

Rajabi-Vandecchali et al. [6] emphasized that ET data is critical for real-time control systems during agricultural operations and for enhancing the efficiency of powered equipment. Janulevičius et al. [7] identified EP as a fundamental parameter for quantitatively assessing fuel consumption and emissions under real-world conditions and optimizing tractor performance and environmental sustainability. Zhu et al. [8] highlighted that the SR is a major determinant of traction efficiency and fuel consumption. Real-time monitoring and regulation of the SR within an optimal range can enhance operational performance and energy efficiency [9]. Kim et al. [10] reported that AP accounts for >70% of the total power consumption during rotary tillage—one of the tractor’s primary tasks—underscoring the need for the precise measurement and analysis of AP in powertrain optimization. These four parameters are not merely performance indicators but are integral to quantifying soil–machine interactions and optimizing tractor operation [11]. However, challenges persist, including the need for high-precision sensors, variability in field conditions, and the complexity of real-time measurement systems.

Beyond mechanical parameters, environmental conditions strongly influence the tractor performance, particularly soil physical characteristics. Since tractors operate primarily on soil, even minor changes in soil properties can significantly affect traction and operational efficiency. Hence, numerous studies have examined how soil physical properties impact the tractor performance. Variables such as soil moisture content (SMC), cone index (CI), and sand proportion (S_p) have been identified as critical influencing factors [12]. Al-Shammary et al. [13] conducted a study to evaluate the performance of a novel digital slippage system for accurately measuring tractor wheel slip across various soil types, including silt, clay, and loam, as well as different tillage methods. Siddique et al. [14] conducted a study to analyze the effect of different gear stages and motor speeds on the traction performance of a single-motor electric tractor in sandy loam soil and confirmed that the L-2500 gear stage achieved the highest traction efficiency and the lowest SR. Janulevičius et al. [15] noted that the SR tends to rise in loamy soils with a CI of ~1.25 MPa, indicating that CI directly affects the SR, traction, and energy efficiency. Ge et al. [16] numerically demonstrated that grouser height acts differently on the shear strength and traction performance of tractor–soil interaction depending on SMC. Min et al. [17] analyzed the effect of soil physical properties (e.g., SMC, CI) on tractor load factor (LF) and reported that regression models incorporating various soil variables could more accurately estimate the LF. Alhassan et al. [18] experimentally measured tractor traction performance during tillage operations in tropical Alfisol soil and developed and validated a multiple linear regression model to estimate draft force (DF), thereby examining the effects of major soil and operational variables on tractor performance. Al-Dosary et al. [19] confirmed that in tillage operations, rear wheel slip of the tractor varies according to SMC and soil composition (clay, clay loam), showing a tendency for slip to decrease as SMC increases. Angelucci et al. [20] reported that tractors exert greater traction but lower efficiency on softer soils, whereas on firmer soils, traction decreases but efficiency improves. These findings have advanced the understanding of soil–tractor interactions and laid the groundwork for developing more efficient tractor operational strategies.

However, much of the existing literature focuses on limited or specific soil conditions, highlighting a gap in research that comprehensively evaluates tractor performance across diverse soil environments. As the relationship between soil physical properties and tractor performance has become clearer, there is a growing interest in estimating tractor performance based on soil conditions.

Before the advent of machine learning, attempts to estimate tractor performance primarily relied on statistical methods such as regression models or simulation techniques. Almaliki et al. [21] developed a simulation program to estimate tractor and implement performances in agricultural fields using CI and tillage depth as input variables. Their model successfully estimated performance metrics such as the DF, SR, and fuel consumption based on variations in CI and tillage depth. Similarly, Seifu et al. [22] proposed a mathematical model to estimate wheel slippage using variables such as the travel speed, tractor weight, SMC, soil density, and soil shear strength. Validation using field experiment data confirmed the model's ability to accurately estimate slippage. Al-Mastawi et al. [23] studied the effects of soil moisture, tire inflation pressure, and travel speed on the tractive performance of a 2WD tractor. They found that the optimal condition—14–17% soil moisture, 85 kPa tire pressure, and 3.15 km/h speed—resulted in the lowest slippage and the highest tractive efficiency (TE). Kim et al. [24] investigated tractor performance in response to CI variations through One-way analysis of variance and least significant difference multiple range tests. Regression analysis was then used to identify the optimal relationship between the dynamic traction ratio (DTR) and TE, revealing that TE increased with increasing CI and reached its maximum when the DTR ranged between 0.45 and 0.55.

More recently, with advancements in artificial intelligence, machine-learning-based estimation techniques have been increasingly employed for agricultural machinery. Machine learning is particularly advantageous in agriculture because it can model nonlinear relationships among variables. Shafaei et al. [25] performed data-mining-based intelligent simulation to estimate the wheel SR using travel speed, tillage depth, and drive type as input variables. Their findings showed that an ANFIS model outperformed an artificial neural network (ANN) in estimating nonlinear slip behavior. Najafi et al. [26] explored the feasibility of estimating ET and fuel consumption using an ANN model, with exhaust gas and lubricant oil temperatures used as input variables. However, studies that quantitatively estimate tractor performance for diverse soil physical properties remain limited. Al-Sager et al. [27] conducted a study to predict the specific fuel consumption of a tractor during tillage operations using an artificial neural network (ANN). They developed an ANN model that incorporated soil conditions, operating speed and depth, and implement width, demonstrating higher prediction accuracy than traditional regression models. Most input variables in the study focused on tractor mechanical parameters, with limited consideration of soil properties, reducing the model's generalizability across diverse field conditions.

Hence, this study aimed to develop a machine learning model that simultaneously estimates key tractor performance indicators—ET, EP, SR, and AP—using various soil physical properties, namely SMC, CI, and sand proportion as input variables. The specific goals of this research were as follows.

- (1) To establish an experimental environment and a data acquisition (DAQ) capable of evaluating soil physical properties and tractor performance and collecting data under various field conditions.
- (2) To analyze the influence of SMC, CI, and sand proportion on ET, EP, SR, and AP.
- (3) To develop and validate a machine-learning-based model that estimates tractor performance for diverse soil physical characteristics using field-collected data.

2. Materials and Methods

2.1. Tractor-Implement System

2.1.1. Agricultural Tractor

In this study, a 78 kW agricultural tractor (S07, Tym Co., Ltd., Gongju, Republic of Korea) widely used in Republic of Korea was used. The tractor measured 4225 mm in length, 2140 mm in width, and 2830 mm in height, with a total weight of 3985 kg. It was equipped

with an in-line, four-cycle engine that delivered a rated torque of 324 Nm at 2300 rpm and a rated power of 78 kW at 2300 rpm. The maximum torque of the engine was 430 Nm at 1400 rpm, with a torque rise of 31.3% [28]. Generally, the rated torque represents the stable power output under standard operating conditions, which is important for fuel efficiency and speed stability. In contrast, the maximum torque at lower engine speed improves traction performance under heavy load conditions by providing higher pulling capacity.

2.1.2. Data Measurement System

The tractor was equipped with a Tier-4-compliant electronic control unit. Three key attributes were selected for analysis among various soil physical properties: SMC, CI, and soil proportion. A soil moisture sensor (TDR350, Spectrum Technology, Aurora, IL, USA) equipped with two 20-cm-long rods was used to measure SMC at a penetration depth of 20 cm [29]. CI was measured using a cone penetrometer (SC900, Spectrum Technology, Aurora, IL, USA) [30]. The average CI value was calculated based on measurements taken at 25-mm intervals up to a depth of 150 mm [31]. Both sensors used are representative equipment widely used in agricultural fields, and the errors presented in the specifications are considered negligible in relation to the analytical results of this study. The specifications of the soil measurement equipment used are provided in Table 1. Soil samples were collected at a depth of 0–150 mm. The soil particle proportion was determined using the hydrometer method. Finally, soil texture classification was performed according to the United States Department of Agriculture soil classification system.

Table 1. Specification of the soil measurement equipment used in this study.

Items	Specification
Soil moisture sensor	Measurement unit: percentage of volumetric water content (VWC) Range: 0% VWC to saturation Accuracy: $\pm 3.0\%$ VWC
Cone penetrometer	Measurement unit: cone index (kPa) Range: 0–45 cm, 0–7000 kPa Accuracy: ± 1.25 cm and ± 103 ka

The measurement system was configured as shown in Figure 1 to enable real-time monitoring of the tractor's engine performance, axle loads, travel speed, and key operating parameters. To account for load differences between the front and rear axles, torque sensors (MW_15kNm, Manner, Germany) were installed on the left and right sides of both the front and rear axles, and rotational speed was measured using a proximity sensor (PRDCMT30-25DO, Autonics, Republic of Korea). Figure 2 shows the installation of the torque and proximity sensors on the tractor. All sensor data were collected in real time and transmitted to the data acquisition (DAQ) system (CRONOS Compact CRC-400-11, IMC, Berlin, Germany) via the Controller Area Network (CAN) channel.

2.2. Data Collection

2.2.1. Field Experiment

This study was conducted using data collected from 10 sites located in Chungcheongnam-do and Gyeonggi-do. Although the sample size may be somewhat limited, the dataset encompasses a diverse range of soil physical properties, providing sufficient representativeness and scope to achieve the research objectives. A total of 60 data samples were collected from each province, resulting in 600 samples overall. Table 2 provides detailed information regarding the size and location of each experimental site. All sites were paddy fields primarily used for rice cultivation, with only stubble remaining during the experiment.

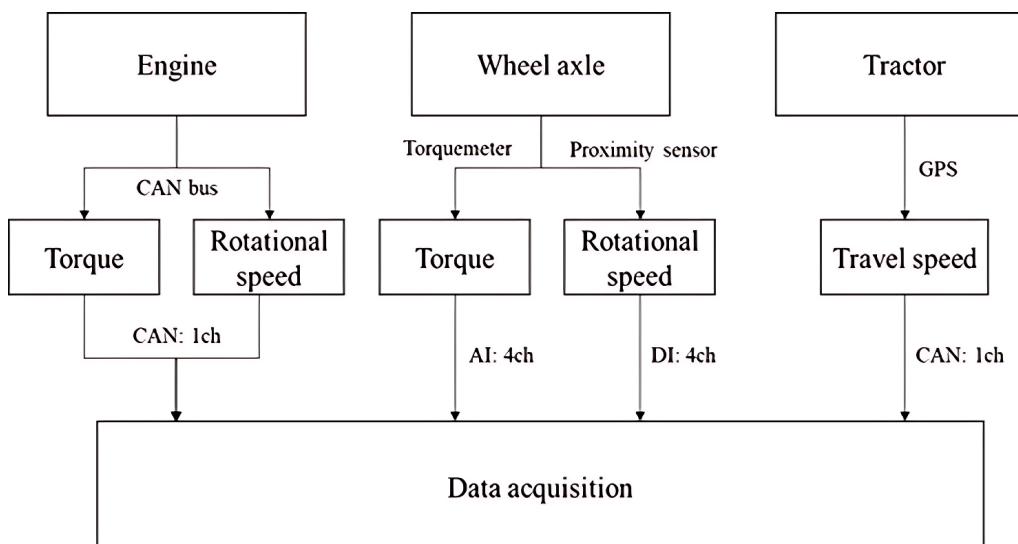


Figure 1. Schematic diagram of data measurement system.

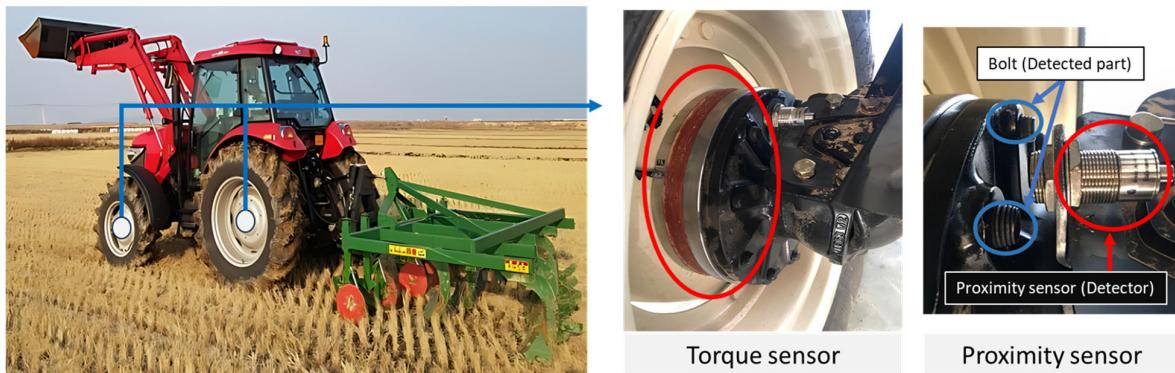


Figure 2. Torque and proximity sensors installed on tractors.

Table 2. Field information of the experiment sites.

Site	Field Size (m^2)	Province	City
S1	6000	Chungcheongnam-do	Seosan
S2	6000	Chungcheongnam-do	Seosan
S3	4000	Chungcheongnam-do	Cheongyang
S4	4000	Chungcheongnam-do	Cheongyang
S5	3000	Gyeonggi-do	Anseong
S6	4000	Gyeonggi-do	Anseong
S7	6000	Chungcheongnam-do	Dangjin
S8	4000	Chungcheongnam-do	Dangjin
S9	4000	Chungcheongnam-do	Dangjin
S10	4000	Chungcheongnam-do	Dangjin

2.2.2. Field-Data-Based Analysis of Soil Physical Properties and Tractor Performance

Figure 3 presents the analysis results of soil physical properties determined at the ten sites. Substantial variability was observed across all sites. Regarding SMC, the greatest variation was recorded at Site 3, ranging from 18.58% to 46.24%. For CI, the largest variation was observed at Site 1, with values ranging from 236.6 to 2085.0 kPa. Overall, CI exhibited a higher range of variation compared to SMC.

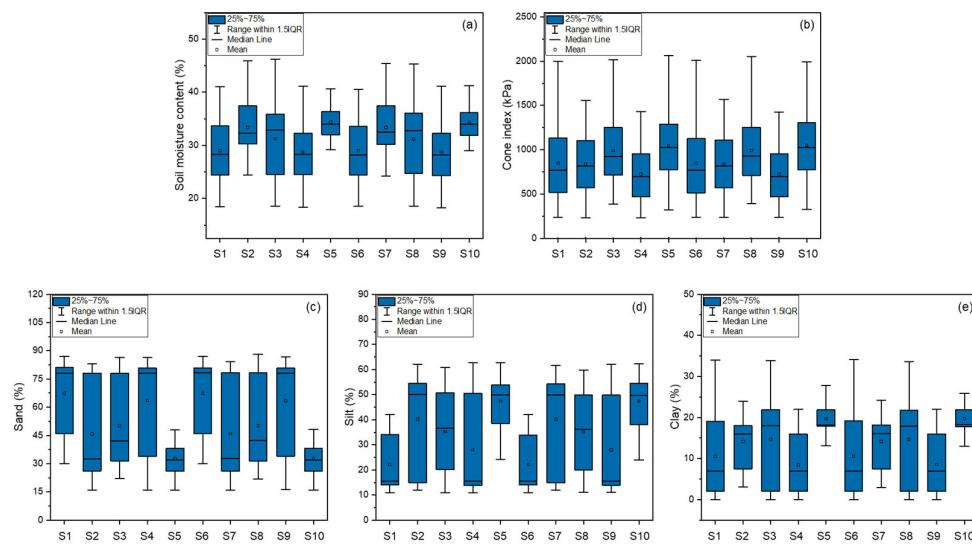


Figure 3. Results of box plot analysis of soil physical property measured on field experiment sites (e.g., S1 refers to site 1): (a) soil moisture content (SMC), (b) cone index (CI), (c) sand, (d) silt, (e) clay.

Regarding soil particle proportion (sand, silt, and clay proportions), the sand proportion was generally higher than the silt and clay sand proportions. The lowest sand proportion was observed at Site 5 (15.91%), while the highest was observed at Site 8 (88.18%). In this analysis, outliers were identified based on the boxplot criterion, where values below $Q1 - 1.5 \times \text{IQR}$ or above $Q3 + 1.5 \times \text{IQR}$ were defined as outliers.

The tractor performance metrics measured at each site are shown in Figure 4. The lowest average ET was recorded at Site 9 (275.58 Nm) and the highest at Site 5 (315.71 Nm). The lowest average engine rotational speed was measured at Site 10 (2267.68 rpm), while Site 9 showed the highest value (2357.77 rpm). The lowest average EP was recorded at Site 9 (67.19 kW) and the highest at Site 10 (75.61 kW). Notably, the box plot patterns for ET and EP were similar, indicating that ET had a more pronounced impact on EP than engine speed (ES). The average SR was the lowest at Site 1 (10.06%) and the highest at Site 5 (14.48%). For AP, the lowest average value was observed at Site 1 (48.42 kW) and the highest at Site 2 (59.15 kW).

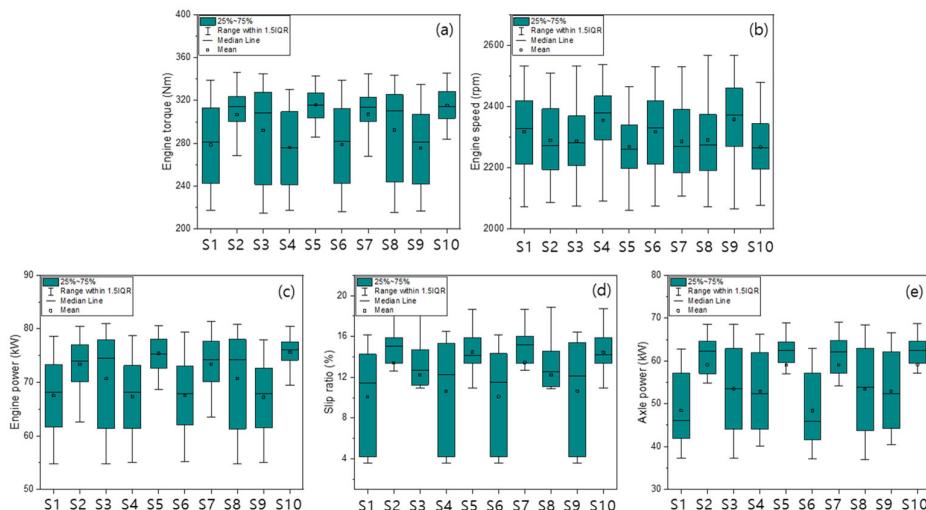


Figure 4. Results of box plot analysis of tractor performance measured on field experiment sites (e.g., S1 refers to site 1): (a) engine torque, (b) engine speed, (c) engine power, (d) slip ratio, (e) axle power.

Table 3 presents the statistical summary of the measured data, with outliers removed based on field observations. SMC and CI ranged from 18.2% to 46.25% and 223 to 2104 kPa, respectively. The CV values for SMC and CI were 18.0% and 43.0%, respectively, indicating considerable variability in CI. The sand, silt, and clay sand proportion ranges were 15.91% to 88.18%, 10.90% to 62.76%, and 0% to 34.11%, respectively. The clay sand proportion exhibited the highest CV (62.8%), reflecting significant spatial heterogeneity in soil texture across the field sites. ET, ES, and EP ranged from 214.6 to 345.8 Nm, 2060.1 to 2567.5 rpm, and 54.75 to 81.32 kW, respectively. These parameters showed low variability, with CV values below 12.6%, suggesting consistent soil resistance or well-maintained operational conditions that minimized abrupt engine performance fluctuations. Meanwhile, SR and AP showed higher variability, lying within ranges of 3.5% to 18.9% and 36.89 to 69.10 kW, respectively, showing CV values of 37.1% and 18.3%, respectively. These variations stemmed from changes in soil properties and working environments. Overall, the soil physical properties exhibited the highest degree of variability among the measured parameters, likely due to the heterogeneous nature of field conditions and regional soil differences.

Table 3. Statistical analysis results of soil physical properties and engine performance indicators.

Items		Max.	Avg.	Std.	CV (%)
Soil physical property	SMC	46.24	31.32	5.64	18
	CI	2104	891	383	43
	Sand	88.18	51.9	24.26	46.7
	Silt	62.76	34.59	17.45	50.4
	Clay	34.11	13.5	8.48	62.8
Engine performance indicators	ET	345.8	293.8	37	12.6
	ES	2567.5	2303.7	122.3	5.3
	EP	81.32	70.87	7.03	9.9
	SR	18.94	12.17	4.52	37.1
	AP	69.1	54.6	9.97	18.3

2.3. Data Analysis

2.3.1. Combination of Input Variables

Table 4 presents the configuration of four models developed to estimate ET, EP, SR, and AP by utilizing different combinations of input variables. The input variables were selected based on their practicality and impact on the estimation accuracy.

Table 4. Combination of variables used for model development.

Site	Field Size (m ²)	Location
S1	6000	Chungcheongnam-do
S2	6000	Chungcheongnam-do
S3	4000	Chungcheongnam-do
S4	4000	Chungcheongnam-do
S5	3000	Gyeonggi-do
S6	4000	Gyeonggi-do
S7	6000	Chungcheongnam-do
S8	4000	Chungcheongnam-do
S9	4000	Chungcheongnam-do
S10	4000	Chungcheongnam-do

- (1) Model A estimated the tractor performance using only SMC because SMC more significantly influences the tractor performance compared to other soil physical properties [18].

- (2) Model B used only CI as the input variable. Since models with a few input variables simplify data collection, models A and B were designed to study the effects of SMC and CI, which are measured in real time using portable sensors [10].
- (3) Model C incorporated SMC and CI as input variables, aiming to improve the estimation accuracy by accurately reflecting soil physical characteristics and tractor performance.
- (4) Model D included all available input variables—SMC, CI, and sand proportion—based on the premise that incorporating more variables enhances the estimation performance.

These models were designed to balance estimation accuracy with model practicality by acknowledging that the combination of input variables directly influences the overall model performance.

2.3.2. Tractor Performance Analysis

EP was calculated based on the measured ET and rotational speed using Equation (1):

$$P_a = \frac{2\pi TN}{60,000} \quad (1)$$

where P_a is the EP requirement (kW), T is the torque (Nm), and N is the engine rotational speed (rpm).

The actual travel speed of the tractor was measured using a GPS device capable of recording positional coordinates at one-second intervals, while the theoretical speed was calculated based on the rotational speed of the tractor's axle. The calculated theoretical speed ranged from 6.01 to 8.43 km/h, and the actual speed ranged from 4.87 to 8.13 km/h. Based on these theoretical and actual speeds, the SR was calculated using Equation (2):

$$s = \left(\frac{V_0 - V_a}{V_0} \right) \times 100 \quad (2)$$

where V_0 is the theoretical speed ($\text{km}\cdot\text{h}^{-1}$) and V_a is the actual travel speed ($\text{km}\cdot\text{h}^{-1}$).

2.3.3. Model Accuracy Evaluation Metric

This study employed three machine learning models—Decision Tree (DT), CatBoost, and LightGBM—to estimate tractor performance. A structural overview of each model is presented in Figure 5. The entire dataset was randomly shuffled and evaluated using K-fold cross-validation to assess the generalization performance of each model. To ensure the reproducibility of the results, all random processes were controlled by setting a fixed random seed of 42.

The performance of the developed estimation models was evaluated using the following statistical metrics: coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and relative deviation (RD). MAPE expresses individual prediction errors as a percentage of the actual values, enabling intuitive interpretation, whereas RD compares RMSE with the mean actual value to provide a relative assessment of the overall error magnitude; therefore, both metrics were used in parallel. These metrics were calculated using Equations (3)–(7), as shown in the following formulas:

$$R^2 = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2 - \sum_{i=1}^N (y_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (4)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (5)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^{i=N} \left| \frac{1}{y_i} (y_i - \hat{y}_i) \right| \times 100(\%) \quad (6)$$

$$RD = \frac{RMSE}{Mean} \times 100(\%) \quad (7)$$

where y_a is the mean measured tractor performance, y_i is the i th measured tractor performance, and \hat{y}_i is the i th estimated tractor performance.

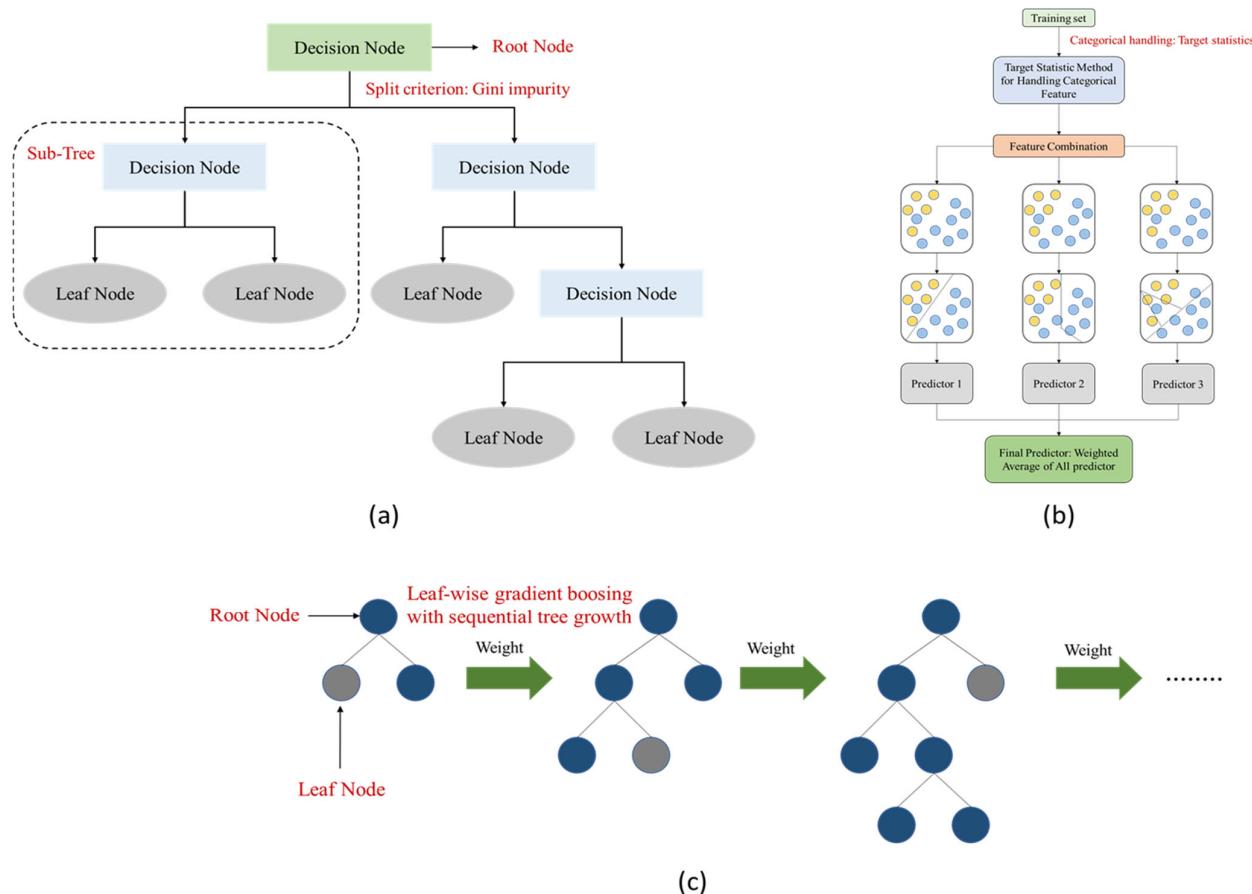


Figure 5. Overview of machine-learning models: (a) Decision Tree (DT), (b) CatBoost, (c) LightGBM.

In general, an R^2 value greater than 0.70 is considered to indicate acceptable performance, while values in the range of 0.80–0.90 or higher are regarded as evidence of strong predictive ability in engineering and agricultural sciences [32]. RMSE is more meaningful when interpreted relative to the data range or mean rather than as an absolute value, and values less than 10% of the mean observed value are considered satisfactory accuracy in studies on agricultural machinery and soil-machine interactions [33].

Field-collected data used in this study were described using specific statistical parameters, including the maximum, minimum, average, standard deviation (SD), and coefficient of variation (CV). The CV was calculated using Equation (8):

$$CV = \frac{SD}{Average} \times 100(\%) \quad (8)$$

Identifying overfitting during model development is critical as overfitting occurs when the model becomes overly tailored to the training data, resulting in poor estimation performance on new, unseen data. To assess overfitting, two additional metrics were used:

1. Train-to-test loss ratio (TTLR): This variable represents the ratio of the mean squared errors (MSE) of the test and training datasets. A TTLR of <1.5 indicates underfitting, while a TTLR of >1.5 indicates overfitting.
2. R² Gap: This measures the difference between the R² values of the training and test datasets. A large R² Gap value suggests a strong likelihood of overfitting, implying the model performs significantly better on training data than test data.

The MSE, TTLR, and R² Gap were determined using Equations (9)–(11).

$$MSE = \frac{1}{N} \sum_{i=1}^{i=N} (y_i - \hat{y}_i)^2 \quad (9)$$

$$TTLR = \frac{MSE_{test}}{MSE_{train}} \quad (10)$$

$$R^2 \text{ Gap} = |R^2_{train} - R^2_{test}| \quad (11)$$

where y_i denotes the i th measured tractor performance and \hat{y}_i denotes the i th estimated tractor performance.

2.3.4. Software

The estimation models were run on a workstation with a 64-bit Windows 11 Pro operating system equipped with a 4.8-GHz processor and 256 GB of memory. Model development was conducted in a Python 3.9-based environment using Google Colab (2025 version). Core machine learning libraries were utilized, including sci-kit, DT, CatBoost, and LightGBM.

2.3.5. Overall Process

Figure 6 presents the overall workflow of this study. First, the input variables were set as SMC, CI, and sand proportion, while the output variables were ET, EP, SR and AP. The data preprocessing stage involved a description of the data distribution and a data cleaning process. The dataset was then split into a training set (70%) and a test set (30%). In the tractor performance prediction model development stage, three machine learning algorithms—DT, CatBoost, and LightGBM—were applied. Hyperparameter optimization was conducted using Optuna in combination with 10-fold cross-validation on the training set. Model performance was evaluated using the test set (30%) based on eight evaluation metrics: R², RD, RMSE, MSE, MAE, TTLR, MAPE, and R² Gap. The best-performing model, CatBoost, was selected based on these evaluation results. For the sensitivity analysis, SHAP interpretation was performed using the best model (CatBoost) and a separate test subset (30%) to analyze the global importance and contribution of each input variable to the prediction of tractor performance.

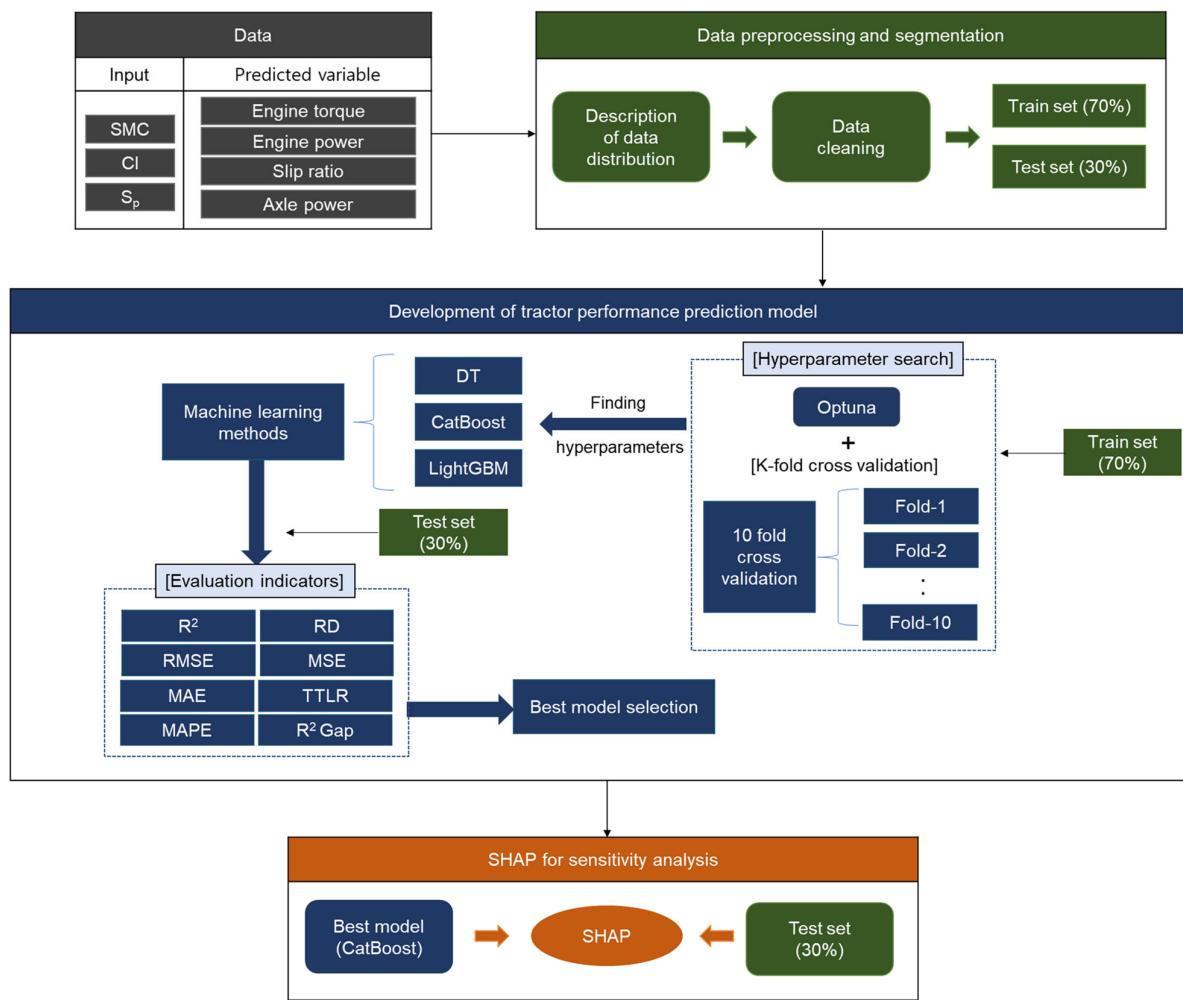


Figure 6. Framework for machine learning-based tractor performance prediction and interpretation.

3. Results

3.1. Effect of Soil Physical Properties on Tractor Performance

Figure 7 presents the correlation matrix between the key soil physical properties and tractor performance metrics. The correlation coefficients ranged from a maximum of 0.80 (between SMC and ET as well as clay and EP) to a minimum of -0.81 (between sand and EP). SMC showed the strongest overall correlation with tractor performance among the soil properties. Specifically, SMC had the greatest influence on ET, with a correlation coefficient of $r = 0.80$. For EP, the most influential properties in the descending order of effect were clay proportion ($r = 0.80$), SMC ($r = 0.79$), silt proportion ($r = 0.74$), CI ($r = 0.39$), and sand proportion ($r = -0.81$). The positive correlation between the clay proportion and EP suggested that as the clay proportion increased soil cohesion, a higher DF was required, which subsequently led to increased EP. SMC showed the highest correlation with SR ($r = 0.74$). Meanwhile, AP was most influenced by silt proportion ($r = 0.68$), followed by SMC ($r = 0.57$), clay proportion ($r = 0.50$), CI ($r = 0.25$), and sand proportion ($r = -0.66$). Correlation coefficients among sand, silt, and clay proportions were $r = 0.71$ (for silt and clay proportions), $r = -0.86$ (clay and sand proportions), and $r = 0.97$ (silt and sand proportions). These values indicate a comprehensive relationship among the particle types, where an increase in one type typically corresponded to a decrease in another. This interplay reflected the inherent regulatory mechanisms of soil physical properties, which influence soil aggregate structure and mechanical behavior.

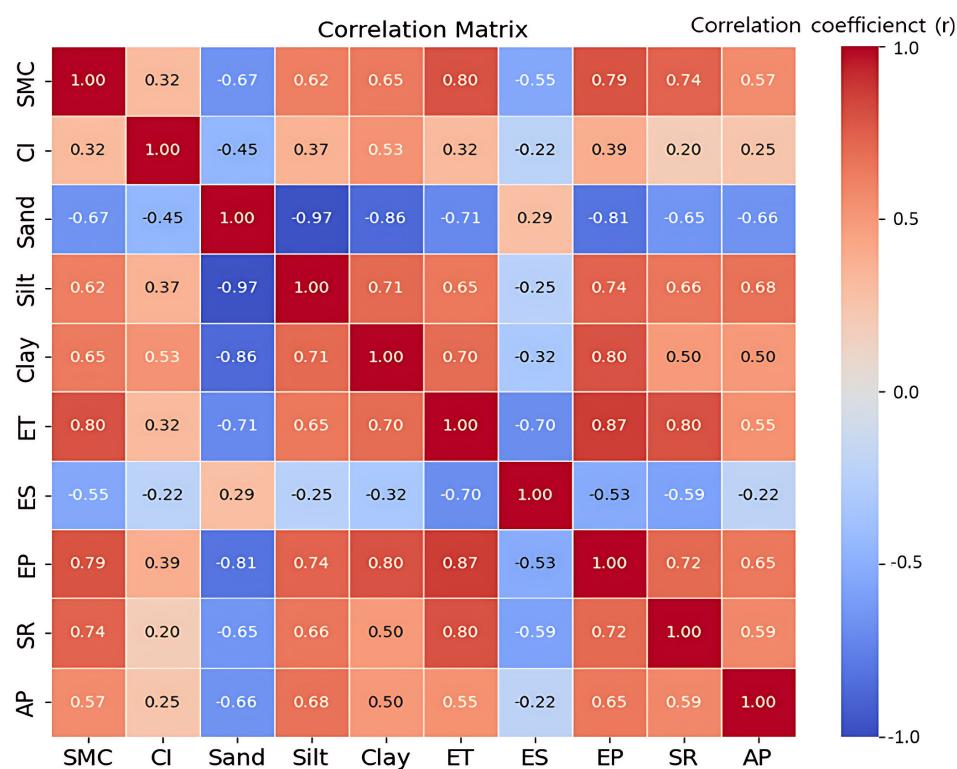


Figure 7. Correlation analysis results between soil physical properties and tractor performance data. Note: SMC: Soil moisture content (%); CI: Cone index (kPa); ET: Engine torque (Nm); ES: Engine speed (rpm); EP: Engine power (kW); SR: Slip ratio (%); AP: Axle power (kW).

Table 5 presents the performance of the developed regression models. The R^2 Adj ranges from 0.104 to 0.771, indicating that each model explains between approximately 10.4% and 77.1% of the variance in tractor performance. Models A and B are univariate regression models that use single soil physical properties—SMC, CI, respectively. Model C is a bivariate model combining these two variables. Model D is a multivariate regression model that additionally incorporates the sand proportion as a soil texture variable, which led to improved estimation accuracy in most cases. For ET, Model D achieved the highest explanatory power with an R^2 Adj of 0.693. Similarly, for EP, Model D showed superior performance with an R^2 Adj of 0.771. These results suggest that including the sand proportion can enhance the accuracy of regression models. In contrast, for SR and AP, the improvement in explanatory power was relatively limited, even with the multivariate model, and the overall prediction accuracy remained low.

Table 6 displays the results of the ANOVA analysis for each regression model, including degrees of freedom (Df), sum of squares (SS), and mean squares (MS). In general, a higher F-value indicates that the regression model more effectively explains the variability in the dependent variable. Model A exhibited a higher F-value than Model B, suggesting stronger explanatory power. In contrast, Models C and D showed relatively lower F-values compared to Model A, indicating weaker explanatory capabilities. These results imply that SMC is more effective than CI in explaining tractor performance. The p -values, calculated based on the F-values and degrees of freedom, represent the probability that there is no significant relationship between the dependent and independent variables. All models demonstrated statistical significance at the $p < 0.01$ level.

Table 5. Regression model for estimating tractor performance.

Output	Model	Regression Model	R ²	R ² Adj	S.E.
Engine torque	A	ET = 5.23SMC + 129.7339	0.637	0.636	22.3
	B	ET = 0.0313CI + 265.8501	0.105	0.104	35.1
	C	ET = 5.07SMC + 0.00768CI + 128.0591	0.643	0.642	22.2
	D	ET = 3.82SMC – 0.000800CI – 0.499Sp + 200.8467	0.695	0.693	20.5
Engine power	A	EP = 0.987SMC + 39.9625	0.627	0.627	4.30
	B	EP = 0.00724CI + 64.4265	0.155	0.154	6.47
	C	EP = 0.924SMC + 0.00292CI + 39.3251	0.650	0.649	4.17
	D	EP = 0.558SMC + 0.000451CI – 0.145Sp + 60.5472	0.773	0.771	4.17
Slip ratio	A	SR = 0.590SMC – 6.3114	0.542	0.542	3.06
	B	SR = 0.00238CI + 10.0487	0.041	0.039	4.43
	C	SR = 0.599SMC – 0.000415CI – 6.2210	0.544	0.542	3.06
	D	SR = 0.441SMC – 0.00148CI – 0.0627Sp + 2.9293	0.599	0.597	2.87
Axe power	A	AP = 1.000SMC + 23.1241	0.323	0.322	8.21
	B	AP = 0.00658CI + 48.7421	0.064	0.062	9.66
	C	AP = 0.960SMC + 0.00210CI + 22.6669	0.329	0.327	8.19
	D	AP = 0.402SMC – 0.00168CI – 0.221Sp + 54.9721	0.470	0.468	7.28

Table 6. ANOVA results for each regression model.

Output	Model	Degrees of Freedom (Df)	Sum of Squares (SS)	Mean Squares (MS)	F-Value	p-Value	Variable	Tolerance	Variance Inflation Factor (VIF)
Engine torque	A	Regression	1	523,869.8	523,869.8	1049.7	0.000 *	SMC	
	A	Residual	9	298,450.1	499.1				
	B	Regression	1	86,400.7	86,400.7	70.2	0.000 *	CI	
	B	Residual	9	735,919.1	1230.6				
	C	Regression	2	528,537.6	264,268.8	537.0	0.000 *	SMC	0.990
	C	Residual	8	293,782.3	492.1			CI	0.990
	D	Regression	3	571,231.4	190,410.5	452.000	0.000 *	SMC	0.550
	D	Residual	7	251,088.4	421.3			CI	0.794
Engine power	A	Regression	1	18,598.7	18,598.7	1006.500	0.000 *	SMC	
	A	Residual	9	11,049.7	18.5				
	B	Regression	1	4602.0	4602.0	109.900	0.000 *	CI	
	B	Residual	9	25,046.3	41.9				
	C	Regression	2	19,274.8	9637.4	554.600	0.000 *	SMC	0.900
	C	Residual	8	10,373.6	17.4			CI	0.900
	D	Regression	3	22,904.1	7634.7	674.700	0.000 *	SMC	0.550
	D	Residual	7	6744.2	11.3			CI	0.794
Slip ratio	A	Regression	1	6648.4	6648.4	708.900	0.000 *	SMC	
	A	Residual	9	5608.6	9.4				
	B	Regression	1	497.9	497.9	25.300	0.000 *	CI	
	B	Residual	9	11,759.1	19.7				
	C	Regression	2	6662.0	3331.0	355.429	0.000 *	SMC	0.900
	C	Residual	8	5594.9	9.4			CI	0.900
	D	Regression	3	7336.7	2445.6	296.200	0.000 *	SMC	0.550
	D	Residual	7	4920.3	8.3			CI	0.794
Axe power	A	Regression	1	19,287.0	19,287.0	285.828	0.000 *	SMC	
	A	Residual	9	40,351.7	67.5				
	B	Regression	1	3801.7	3801.7	40.715	0.000 *	CI	
	B	Residual	9	55,837.0	93.4				
	C	Regression	2	19,634.9	9817.5	146.512	0.000 *	SMC	0.900
	C	Residual	8	40,003.8	67.0			CI	0.900
	D	Regression	3	28,044.9	9348.3	176.400	0.000 *	SMC	0.550
	D	Residual	7	31,593.8	53.0			CI	0.794

* Significant at $p < 0.01$.

The variance inflation factor (VIF) was used to diagnose multicollinearity, calculated as the reciprocal of tolerance. A VIF value closer to 1 indicates little to no multicollinearity among independent variables, while a value of 10 or higher suggests potential multicollinearity issues. In this study, the VIF values for all models ranged from 1.111 to 2.059, confirming that none of the models exhibited multicollinearity problems.

3.2. Machine Learning Model-Based Estimation of Tractor Performance

3.2.1. Engine Torque

Figure 8 illustrates the results of ET estimation using different combinations of the input variables. The analysis revealed that Models C and D showed high estimation performance, with data points closely aligning with the reference line, whereas Models A and B exhibited greater data dispersion and relatively lower predictive accuracy. In addition, similar distributions were observed in both the training and test datasets, suggesting that the risk of overfitting was not significant. However, some deviations were noted in the high engine torque range, which may be attributed to the models' limited ability to fully capture the nonlinearities of soil-machine interactions under heavy-load conditions.

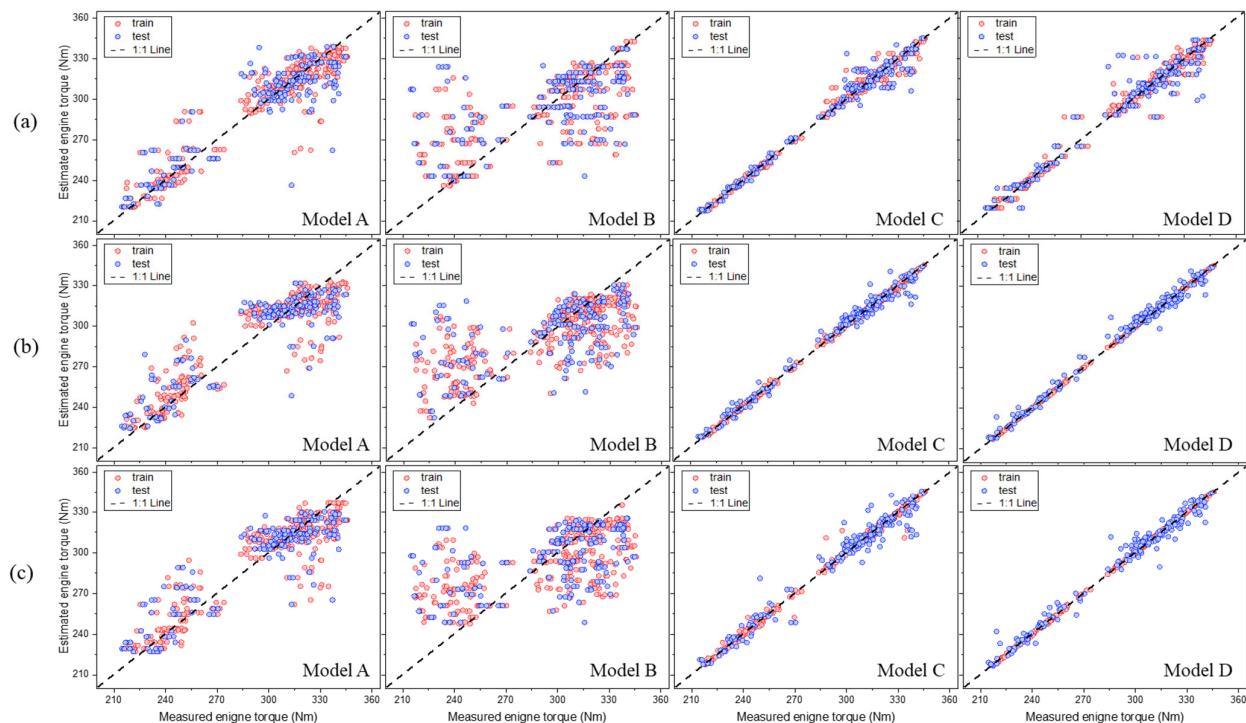


Figure 8. Scatter plots of the estimated and estimated values for the engine torque: (a) DT, (b) CatBoost, and (c) LightGBM. Note: The 1:1 reference line represents perfect agreement between the observed and predicted values. The closer the data points are to this line, the better the predictive performance of the model is considered to be.

Table 7 summarizes the performance of the three machine learning algorithms—DT, CatBoost, and LightGBM—in estimating ET. The models were evaluated using five metrics: R^2 , RMSE, MAE, MAPE, and RD. Model robustness and overfitting were assessed using the R^2 Gap and TTLR metrics. Across all models, the R^2 values ranged from 0.220 to 0.990, the RMSE ranged from 1.34 to 9.15 Nm, the MAE ranged from 1.34 to 9.46 Nm, MAPE ranged from 0.64% to 9.17, and the RD ranged from 1.24% to 11.06%. Model B exhibited the lowest R^2 value, indicating limited estimation performance. In contrast, CatBoost-based model D achieved the highest estimation accuracy with R^2 of 0.990, RMSE of 3.65 Nm, MAE of 2.80 Nm, MAPE of 0.96%, and RD of 1.24%.

Table 7. Performance evaluation of the machine learning models in estimating engine torque.

Items	R ²		R ² Gap	RMSE (Nm)		TTLR	MAE (Nm)		MAPE (%)		RD (%)	
	Train	Test		Train	Test		Train	Test	Train	Test	Train	Test
A	DT	0.834	0.747	0.088	5.17	6.31	1.49	5.82	6.96	3.69	4.36	5.17
	CatBoost	0.883	0.780	0.103	5.54	6.78	1.50	8.70	9.46	2.98	4.22	4.27
	LightGBM	0.822	0.703	0.119	5.73	7.82	1.86	6.47	7.33	3.95	4.83	5.36
B	DT	0.411	0.220	0.192	8.60	9.15	1.13	5.55	6.67	7.44	8.18	9.74
	CatBoost	0.532	0.347	0.185	5.51	6.59	1.43	5.90	6.71	7.21	8.54	8.69
	LightGBM	0.424	0.300	0.124	4.30	5.55	1.66	6.40	6.43	8.10	9.17	9.64
C	DT	0.963	0.848	0.115	7.18	8.20	1.31	4.67	8.00	1.60	2.70	2.45
	CatBoost	0.999	0.972	0.027	1.15	1.34	1.37	0.93	1.34	1.32	1.47	1.39
	LightGBM	0.991	0.949	0.043	3.45	4.23	1.50	2.06	2.51	0.72	1.87	1.17
D	DT	0.961	0.915	0.046	7.28	8.06	1.23	3.85	6.23	1.32	2.10	2.48
	CatBoost	0.999	0.990	0.009	3.30	3.65	1.22	2.25	2.80	0.86	0.96	1.10
	LightGBM	0.999	0.968	0.031	5.32	5.65	1.13	3.72	4.68	0.26	0.64	1.45

3.2.2. Engine Power

Figure 9 summarizes the performance metrics of the EP estimation models, showing the comparison of the measured and estimated EP values across different combinations of the input variables. The results demonstrated that alignment with the 1:1 reference line improved as the number of input variables increased. Models C and D exhibited data points that closely aligned with the 1:1 reference line across the entire output range, demonstrating high estimation performance and stability. In contrast, Models A and B showed instances of overestimation or underestimation in certain ranges. Additionally, Models C and D maintained a high level of consistency between the training and test datasets across the full output range, indicating that the risk of overfitting was not substantial.

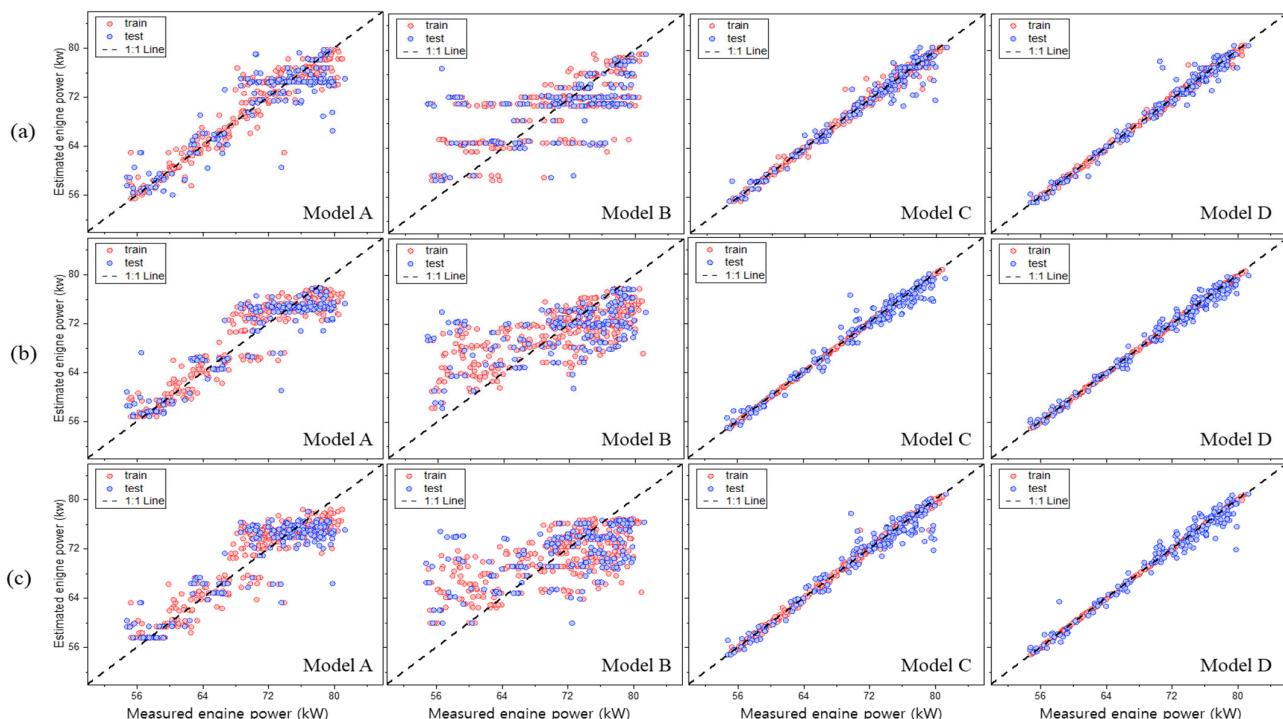


Figure 9. Scatter plots of the estimated and estimated values for the engine power: (a) DT, (b) CatBoost, and (c) LightGBM. Note: The 1:1 reference line represents perfect agreement between the observed and predicted values. The closer the data points are to this line, the better the predictive performance of the model is considered to be.

Table 8 summarizes the performance metrics of the EP estimation models. The analysis revealed that all models, except model B, achieved R^2 values of at least 0.728. Model A achieved an average R^2 of 0.738. Meanwhile, model B exhibited the lowest performance, with an average R^2 of 0.394. Model C showed a significantly improved average R^2 of 0.932 compared to models A and B. When sand proportion was included in model D, the average R^2 was 0.948. These findings indicate that using multiple input variables enhances the estimation performance compared to using a single variable. The overfitting assessment results showed that the R^2 Gap was below 0.181 and the TTLR was below 1.94. However, the TTLR value for Model B was relatively higher compared to the other models, suggesting that when using Model B, techniques such as dropout, early stopping, and cross-validation should be applied to prevent overfitting.

Table 8. Performance evaluation of the machine learning models in estimating engine power.

Items	R^2		R^2 Gap	RMSE (kW)		TTLR	MAE (kW)		MAPE (%)		RD (%)	
	Train	Test		Train	Test		Train	Test	Train	Test	Train	Test
A	DT	0.867	0.728	0.139	2.50	2.87	1.32	1.82	3.01	2.54	4.25	3.53
	CatBoost	0.851	0.749	0.102	2.71	3.53	1.70	2.16	2.77	3.06	3.89	3.83
	LightGBM	0.819	0.737	0.082	2.99	3.61	1.46	2.34	2.84	3.32	3.99	4.22
B	DT	0.591	0.410	0.181	4.46	5.34	1.43	2.88	4.51	4.23	6.60	6.29
	CatBoost	0.562	0.394	0.168	4.53	5.78	1.63	3.50	4.50	5.15	6.60	6.40
	LightGBM	0.499	0.377	0.123	4.94	6.47	1.72	3.74	5.12	5.47	7.43	6.96
C	DT	0.950	0.877	0.073	1.62	2.25	1.94	1.02	1.57	1.42	2.14	2.30
	CatBoost	0.999	0.967	0.032	1.09	1.28	1.39	0.68	0.89	0.97	1.24	0.13
	LightGBM	0.996	0.952	0.045	1.43	1.49	1.08	0.29	0.98	0.42	1.37	1.61
D	DT	0.962	0.898	0.064	1.34	1.49	1.23	0.85	1.37	1.17	1.96	1.88
	CatBoost	0.999	0.980	0.019	0.99	1.00	1.01	0.77	0.72	0.91	1.04	1.13
	LightGBM	0.999	0.965	0.034	1.21	1.38	1.30	1.16	1.99	1.23	1.40	1.30

3.2.3. Slip Ratio

Figure 10 presents the results of the SR estimation models, comparing the measured and estimated SR values for each combination of input variables. Among the developed models, models C and D showed the closest alignment with the 1:1 reference line, indicating the highest estimation accuracy. In contrast, Models A and B exhibited greater data dispersion and showed a clear deviation from the reference line overall, resulting in relatively lower predictive performance. The errors were particularly pronounced in the low- and high-slip regions, which can be attributed to the inability of simple models to adequately capture the nonlinear nature of slip ratio influenced by multiple factors such as soil conditions, traction, and load.

Table 9 summarizes the performance metrics of each model. In particular, Model B, which utilized the CatBoost algorithm, achieved the highest accuracy with an R^2 of 0.983, an RMSE of 0.61%, an MAE of 0.35%, a MAPE of 3.84%, and an RD of 2.97%. Although the MAPE for SR prediction was relatively high, this was due to the disproportionate influence of relative errors in sections where the actual values were small. Given that the absolute error metrics (RMSE and MAE) were very low, the prediction reliability is still considered high.

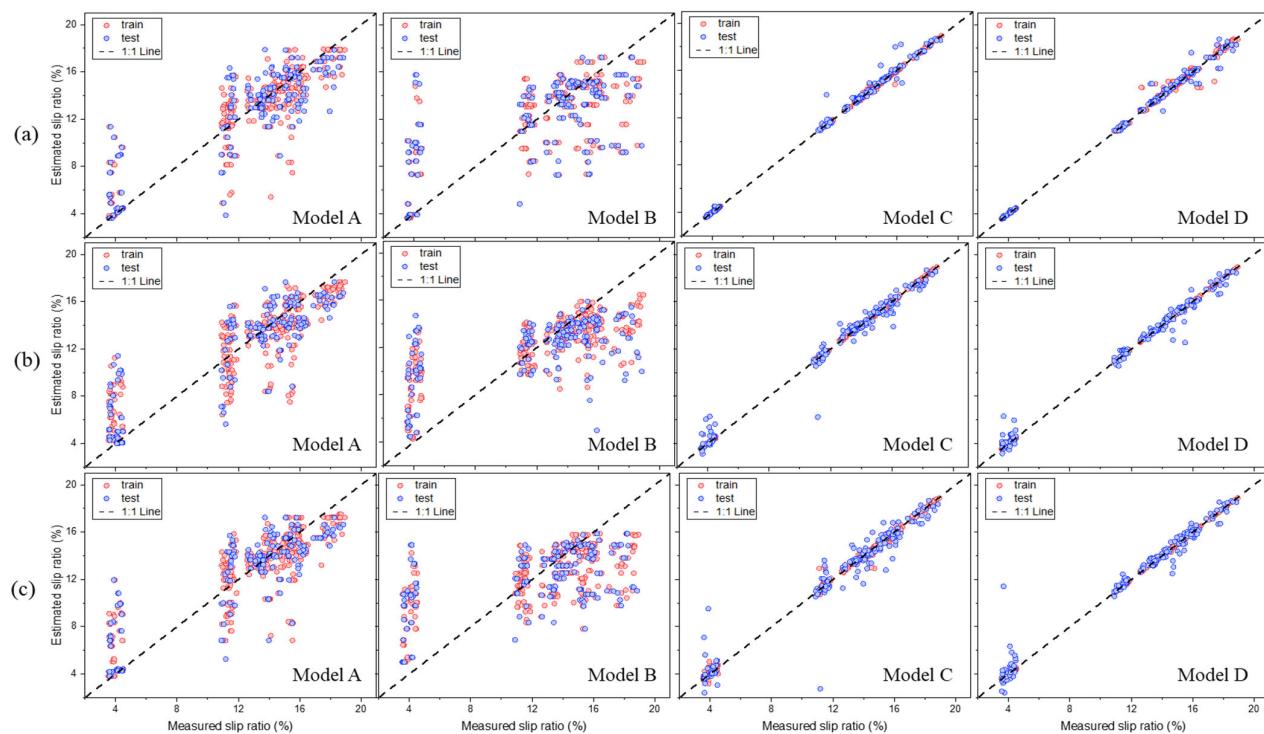


Figure 10. Scatter plots of the estimated and estimated values for the slip ratio: (a) DT, (b) CatBoost, and (c) LightGBM. Note: The 1:1 reference line represents perfect agreement between the observed and predicted values. The closer the data points are to this line, the better the predictive performance of the model is considered to be.

Table 9. Performance evaluation of the machine learning models in estimating the slip ratio.

Items	R ²		R ² Gap	RMSE (%)		TTLR	MAE (%)		MAPE (%)		RD (%)	
	Train	Test		Train	Test		Train	Test	Train	Test	Train	Test
A	DT	0.702	0.562	0.140	2.48	2.96	1.42	1.79	2.30	6.72	7.91	6.50
	CatBoost	0.781	0.642	0.139	2.12	2.67	1.58	1.59	2.08	7.61	7.94	7.58
	LightGBM	0.808	0.635	0.173	1.99	2.70	1.84	1.37	2.01	6.53	6.98	6.47
B	DT	0.534	0.462	0.072	3.11	4.04	1.69	2.07	2.77	8.18	8.70	6.88
	CatBoost	0.583	0.327	0.257	2.94	3.62	1.52	2.10	2.59	6.78	7.19	5.47
	LightGBM	0.485	0.296	0.189	3.27	3.70	1.28	2.40	2.71	6.78	6.57	7.20
C	DT	0.938	0.822	0.116	1.11	1.34	1.45	0.42	0.76	4.86	5.20	3.09
	CatBoost	0.999	0.977	0.022	0.57	0.70	1.53	0.36	0.38	4.47	4.65	4.46
	LightGBM	0.997	0.928	0.069	1.24	1.24	1.00	0.54	0.63	1.66	2.53	1.96
D	DT	0.997	0.919	0.078	1.25	1.34	1.14	0.11	0.46	3.92	4.88	2.06
	CatBoost	0.999	0.983	0.016	0.52	0.61	1.38	0.22	0.35	3.08	3.84	2.18
	LightGBM	0.998	0.955	0.044	0.98	1.00	1.04	0.51	0.54	4.46	4.96	1.45

3.2.4. Axle Power

Figure 11 presents the results of the AP estimation models, comparing the measured and estimated AP values for each combination of input variables. Among the developed models, Model D showed the highest agreement with the 1:1 reference line. In addition, a general trend was observed in which the agreement with the reference line improved as the number of input variables increased. In the AP prediction, several data points in Model C were observed to deviate noticeably from the reference line. This indicates that using only SMC and CI has limitations in predicting AP, and that performance improves when soil texture is included as an additional input variable.

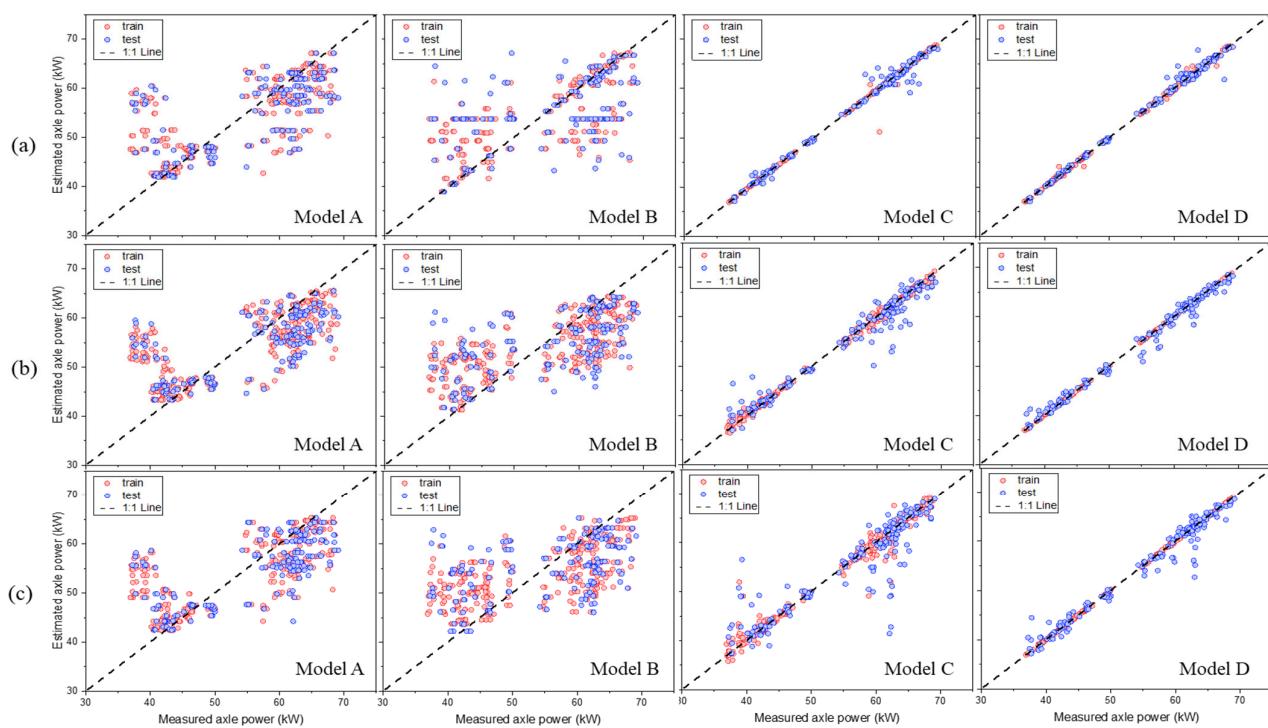


Figure 11. Scatter plots of the estimated and estimated values for the axle power: (a) DT, (b) CatBoost, and (c) LightGBM. Note: The 1:1 reference line represents perfect agreement between the observed and predicted values. The closer the data points are to this line, the better the predictive performance of the model is considered to be.

Table 10 summarizes the performance metrics of the AP estimation models. The analysis showed that the R^2 values ranged from 0.120 to 0.959, RMSE values from 1.27 to 9.46 kW, MAE values from 0.88 to 7.98 kW, MAPE values from 1.55% to 9.30%, and RD values from 0.92% to 6.81%. Model B exhibited lower estimation performance compared to the other models, as indicated by its lower R^2 value and higher RMSE, MAE, MAPE, and RD values. This suggests that using only CI as an input variable reduces estimation accuracy. Relying on a single input variable limits the model's ability to capture the complex and variable conditions encountered in actual field operations. Therefore, incorporating multiple soil-related variables can provide a more comprehensive understanding of field conditions and improve the model's estimation performance.

Table 10. Performance evaluation of the machine learning models in estimating axle power.

Items	R^2		R^2 Gap	RMSE (kW)		TTLR	MAE (kW)		MAPE (%)		RD (%)	
	Train	Test		Train	Test		Train	Test	Train	Test	Train	Test
A	DT	0.424	0.388	0.036	7.61	7.69	1.02	5.56	5.73	6.24	6.39	3.98
	CatBoost	0.537	0.410	0.127	6.82	7.55	1.23	4.99	5.62	6.08	6.11	4.53
	LightGBM	0.621	0.502	0.120	6.17	8.21	1.77	4.30	5.76	8.63	9.30	4.34
B	DT	0.240	0.120	0.120	8.71	9.28	1.14	7.48	7.98	4.81	5.28	6.03
	CatBoost	0.535	0.310	0.224	6.82	8.84	1.68	5.66	7.11	5.19	5.67	4.54
	LightGBM	0.423	0.386	0.036	7.59	9.46	1.55	6.26	7.69	4.39	4.84	3.97
C	DT	0.849	0.743	0.106	3.88	4.73	1.48	1.68	3.00	3.39	2.68	5.07
	CatBoost	0.999	0.889	0.110	3.21	3.31	1.07	0.17	1.52	0.31	3.26	2.38
	LightGBM	0.989	0.831	0.158	4.07	4.09	1.01	1.57	2.29	1.11	1.55	1.95
D	DT	0.958	0.840	0.118	2.06	2.28	1.23	0.93	1.67	1.87	3.68	3.76
	CatBoost	0.999	0.959	0.040	1.11	1.27	1.32	0.79	0.88	1.17	1.87	3.20
	LightGBM	0.999	0.924	0.075	2.21	2.68	1.47	0.14	1.28	2.28	2.67	0.38

3.3. Combination and Evaluation of the Machine Learning-Based Estimation Models

Figure 12 presents the performance comparison of machine learning models for estimating tractor performance using R^2 as the evaluation metric. For ET, Model B recorded the lowest R^2 value, with DT and LightGBM showing particularly low values of 0.220 and 0.300, respectively. In contrast, Model C showed a substantial improvement in R^2 values across all algorithms, with CatBoost achieving the highest value of 0.972. For EP, although Model B exhibited low R^2 values, the other models demonstrated stable and reliable performance, with R^2 values ranging from 0.728 to 0.980. Notably, in Model D, DT, CatBoost, and LightGBM achieved high accuracies of 0.898, 0.980, and 0.965, respectively. For SR, the R^2 values ranged from 0.296 to 0.983, indicating considerable performance differences among the models. Model D achieved the highest estimation accuracy, with DT, CatBoost, and LightGBM recording R^2 values of 0.919, 0.983, and 0.955, respectively. For AP, Model A showed lower accuracy compared to the other models, as using only SMC as an input variable was insufficient to capture the nonlinear characteristics of AP. In contrast, Model C achieved a significant improvement in estimation accuracy by incorporating two soil physical properties—SMC and CI—as input variables. This indicates that including a more diverse set of input variables contributes to improved estimation performance.

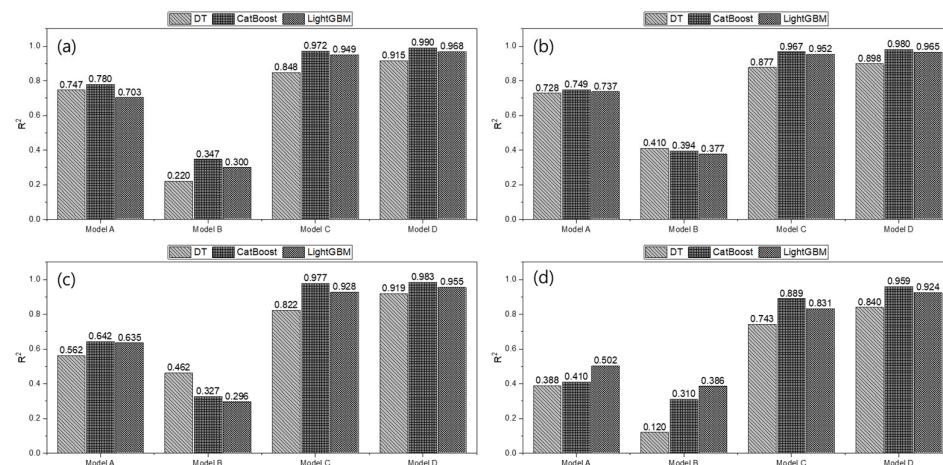


Figure 12. Performance comparison of three machine learning models using R^2 for predicting (a) engine torque, (b) engine power, (c) slip ratio, (d) axle power.

Overall, Model D utilizing CatBoost achieved the highest R^2 values across ET, EP, SR, and AP estimations. This suggests that CatBoost, with its capability to effectively capture nonlinearity and interactions among variables, is well suited for addressing complex problems such as predicting agricultural machinery performance.

Figure 13 presents the comparison of prediction performance for three machine learning models in estimating ET, EP, SR, and AP using the MAPE metric. For ET, CatBoost and LightGBM in Model D achieved very low MAPE values of 0.96% and 0.64%, respectively, demonstrating the best performance. In contrast, Models A and B recorded MAPE values exceeding 4%, indicating that errors increased substantially when input variables were limited. For EP, all three algorithms (DT, CatBoost, and LightGBM) in Model B showed high MAPE values above 6.6%, whereas in Model D, the errors sharply decreased to below 2.0%. This demonstrates that prediction stability improves significantly when sufficient input variables are utilized. For SR, the performance differences among the models were the most pronounced. Models A and B exhibited high MAPE values in the range of 7–9%, whereas Model D substantially improved performance, with CatBoost and LightGBM achieving 3.84% and 4.96%, respectively. For AP, DT in Model A recorded the largest error at 9.30%,

while CatBoost and LightGBM in Model D achieved very low errors of 1.87% and 2.67%, respectively, demonstrating stable performance.

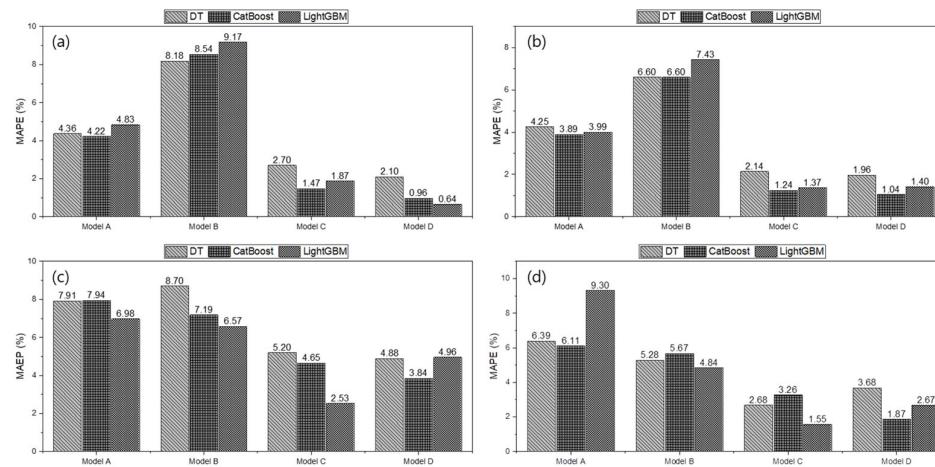


Figure 13. Performance comparison of three machine learning models using MAPE for predicting (a) engine torque, (b) engine power, (c) slip ratio, (d) axle power.

3.4. Hyperparameter Optimization Results

Hyperparameter optimization was conducted using Optuna, focusing on key parameters that influence model complexity and learning behavior for each algorithm. For the DT model, the search space included max_depth ranging from 2 to 8, min_samples_split from 2 to 10, and min_samples_leaf from 1 to 5. The optimal configuration identified was max_depth = 8, min_samples_split = 8, and min_samples_leaf = 2. In the case of the CatBoost model, iterations were explored from 100 to 1000, learning_rate from 0.01 to 0.3, and depth from 3 to 10. The best-performing configuration was found to be iterations = 609, learning_rate = 0.160, and depth = 10. For the LightGBM model, the ranges considered were num_leaves between 20 and 100, learning_rate between 0.01 and 0.3, and n_estimators between 100 and 1000, with the optimal parameters determined as num_leaves = 75, learning_rate = 0.108, and n_estimators = 746. These optimal values were identified through Bayesian optimization using the Tree-Structured Parzen Estimator and contributed to improved model performance [34]. The summary of the results is presented in Table 11.

Table 11. Selected hyperparameter combinations for the machine learning models used to estimate the tractor performance.

DT	CatBoost	LightGBM
max_depth: 8	number of iterations: 609	num_leaves: 75
min_samples_split: 8	learning_rate: 0.160	learning_rate: 0.108
min_samples_leaf: 2	depth: 10	n_estimators: 746

3.5. Shap Interpretation

To identify the key input features influencing the model's prediction results and to enhance the interpretability and reliability of the model, SHAP sensitivity analysis was conducted. In this study, the analysis focused on global interpretation within the SHAP framework. SHAP analysis is based on an additive feature attribution method, where the explanation model $g(z')$ is represented as a linear function of binary features, as defined using Equation (12) [35].

$$g(z') = \emptyset_0 + \sum_{i=1}^M \emptyset_i z'_i \quad (12)$$

where M denotes the number of input features and \emptyset_i represents the Shapley value of feature i , the computation is outlined as follows. Let S denote the set of all features, and S_{-i} denote the set of features excluding feature i . $P(S)$ refers to the power set of S , containing all possible subsets of S , while $|A|$ represents the cardinality, or the number of elements, in set A . The Shapley value (\emptyset_i) for a particular feature i is calculated using Equation (13):

$$\emptyset_i = \sum_{A \in P(S_{-i})} \left[\frac{|A|! \cdot (|F| \cdot |A| - 1)!}{|F|!} \right] [f(A \cup \{i\}) - f(A)] \quad (13)$$

In this process, the sum is taken over all possible subsets A that exclude feature i , with F denoting the set of all features. The terms $f(A \cup i)$ and $f(A)$ represent the model's output with and without the inclusion of feature i in subset A , respectively. This formula quantifies the marginal contribution of feature i across all possible combinations of features, offering a comprehensive view of the possible feature subsets.

Figure 14 presents the results of the sensitivity analysis conducted using the CatBoost-based prediction model. The relative importance of the three input variables—SMC, CI, and Sand—was evaluated for each prediction target based on SHAP values. For ET prediction, the SHAP values for SMC, CI, and Sand were 37.56, 33.26, and 29.18, respectively. SMC showed approximately 28.7% higher importance than Sand, indicating that soil moisture is the most influential factor in estimating ET. In the case of EP prediction, SMC had the highest SHAP value of 3.10, which was 17.0% and 162.7% greater than those of Sand and CI, respectively. This result emphasizes the dominant contribution of SMC to EP estimation. For SR prediction, SMC again had the greatest influence with a SHAP value of 1.92, followed by Sand (1.55) and CI (0.78). The importance of SMC was 23.9% higher than that of Sand, while CI had the lowest impact. In contrast, for AP prediction, Sand was the most important variable with a SHAP value of 6.02, which was 113.5% greater than that of SMC (2.82). This suggests that sand proportion plays a key role in predicting AP. The differences in feature importance among the performance indicators stem from the distinct physical mechanisms that each indicator reflects. ET, EP, and SR were all found to be highly sensitive to variations in soil moisture content, as soil moisture directly influences soil mechanical properties such as shear strength, cohesion, and internal friction angle. An increase in moisture reduces tire–soil friction, leading to higher slip ratios and consequently requiring the engine to deliver greater torque and power. In contrast, AP is largely governed by soil texture and shear resistance, with sand proportion playing a particularly critical role, as it alters soil drainage and strength characteristics, thereby directly affecting power transmission efficiency. Thus, while ET, EP, and SR are closely associated with soil moisture, AP is more strongly influenced by sand content.

Figure 15 presents the SHAP summary plots of the CatBoost model, illustrating the contribution patterns of SMC, Sand, and CI for each prediction target. In the ET prediction, the SHAP values of SMC are broadly distributed and predominantly positive, indicating a tendency for higher soil moisture to increase the predicted values. Sand and CI exhibited both positive and negative effects, but their distributions were narrower compared to that of SMC. For EP and SR predictions, SMC also showed an increasing trend in SHAP values as feature values rose, suggesting a major contribution to the output. CI showed SHAP values largely concentrated around zero, indicating relatively consistent effects across samples. In the case of AP prediction, Sand displayed a wide range of SHAP values with a distinct positive pattern across various feature values, while SMC and CI showed more limited distributions within a narrower range.

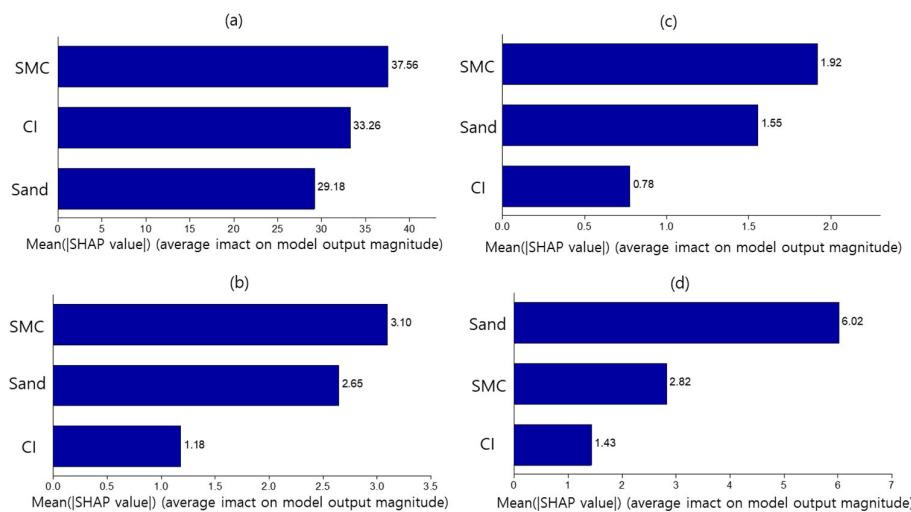


Figure 14. Feature importance of the three different models based on mean absolute SHAP value: (a) engine torque, (b) engine power, (c) slip ratio, (d) axle power.

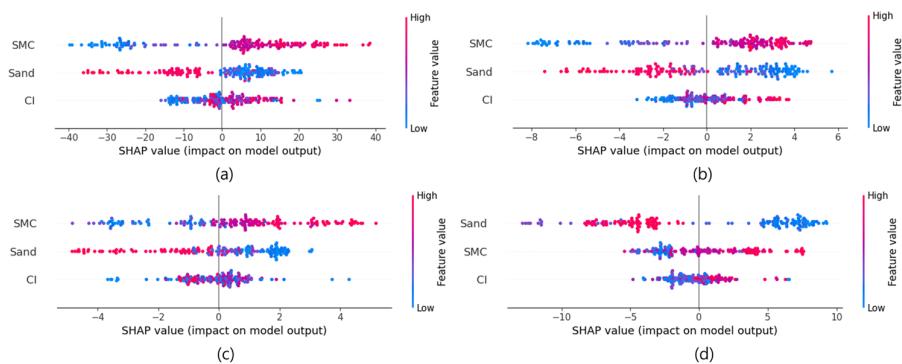


Figure 15. SHAP summary plot of the CatBoost model: (a) engine torque, (b) engine power, (c) slip ratio, (d) axle power.

Overall, the SHAP analysis results revealed patterns that align well with the known positive or negative correlations between each input variable and the corresponding prediction targets.

4. Discussion

In this study, machine learning-based models were developed to estimate key tractor performance indicators using various soil physical properties as input variables. Three algorithms—DT, CatBoost, and LightGBM—were applied, and model performance was evaluated across four input-variable combinations.

Overall, the results indicated that increasing the number of input variables generally improved estimation accuracy. Model D achieved the highest estimation accuracy across all performance indicators, with R^2 values ranging from 0.840 to 0.990. In particular, CatBoost consistently recorded the highest R^2 values among all algorithms, demonstrating its effectiveness in addressing complex estimation problems in agricultural environments. This can be attributed to CatBoost's ability to effectively capture nonlinear relationships and interactions between variables, as well as its capabilities in handling categorical variables, automatic splitting, and built-in regularization techniques, which help prevent overfitting and maintain stable, high performance even with complex agricultural data [36]. LightGBM also demonstrated strengths in stable performance and fast training speed. In particular, compared with Random Forest, CatBoost and LightGBM are advantageous in terms of

learning efficiency and memory usage, while compared with XGBoost, they offer benefits in automated handling of categorical variables and faster computational speed [37].

However, increasing the number of input variables also raised the complexity and computational cost of estimation. Model C, which used only a small number of input variables based on sensors capable of real-time measurement, demonstrated high performance, suggesting that it could serve as a practical and realistic solution for field applications. In contrast, Model B recorded the lowest performance, with R^2 values ranging from 0.120 to 0.462, and showed particularly poor results in estimating SR and AP. This indicates that using CI alone as an input variable is insufficient to capture the complexity of soil–tractor interactions.

Among the three machine learning algorithms, CatBoost achieved the best overall estimation performance. DT offered advantages in interpretability and low computational cost but showed lower estimation accuracy. Overfitting was assessed using the R^2 Gap and TTLR metrics, which ranged from 0.009 to 0.224 and from 1.00 to 1.94, respectively. These values indicate that, with few exceptions, most models achieved generalized estimation performance without significant overfitting.

This study addresses several limitations of previous research. Unlike earlier studies that primarily focused on SR or were conducted under restricted soil conditions, this work simultaneously estimated four key performance indicators—ET, EP, SR, and AP—thereby providing a more comprehensive framework for evaluating tractor performance. In terms of input variables, whereas prior studies often relied on a limited set, such as CI or SMC, this study proposed a multivariate estimation model that incorporates a wider range of soil physical properties, including soil texture composition. Nevertheless, some limitations remain, as data collection was confined to specific conditions, and the increased model complexity associated with additional input variables may lead to efficiency issues.

Future research should apply hyperparameter optimization and regularization techniques to further reduce overfitting. In addition, expanding data collection to include various soil types, such as clay, sandy loam, and silt loam, as well as diverse operating conditions, including plowing, rotary tillage, and harvesting, under different soil moisture levels and seasonal climatic variations, will further enhance the robustness and generalizability of the model. Furthermore, it is necessary to validate the model's performance under extreme soil conditions, such as severe drought or excessive moisture, to more reliably ensure its adaptability in real agricultural environments.

5. Conclusions

In this study, a machine learning model was developed to predict key tractor performance indicators (ET, EP, SR, AP) based on soil physical properties. As a result, the CatBoost model demonstrated the best performance, recording an R^2 value of 0.990 in ET prediction. This goes beyond previous studies that primarily relied on single variables, highlighting the academic significance of integrating multiple soil variables to improve the generalizability and reliability of tractor performance prediction. To develop such a model, it is necessary to first select test fields and measure various soil conditions, then choose a model appropriate for those conditions to predict the tractor's key performance indicators. Furthermore, for practical implementation in the field, a systematic process of collecting soil condition data to be used as input variables is required. However, since this study was based on data collected from a specific region and limited soil conditions, there are inherent limitations in generalizing the findings to various soil types, climates, and working environments. Therefore, future research will focus on expanding the range and diversity of data and validating the model under diverse operating conditions to further strengthen its reliability and applicability.

Author Contributions: Conceptualization, Y.-J.K. and W.-S.K.; Methodology, S.-M.B. and S.-Y.B.; Software, S.-M.B. and S.-Y.B.; Validation, W.-S.K.; Formal analysis, S.-Y.G., S.-M.B. and S.-Y.B.; Investigation, S.-M.B. and S.-Y.B.; Resources, S.-M.B., S.-Y.B. and W.-S.K.; Data curation, S.-M.B., S.-Y.B. and W.-S.K.; Writing—original draft preparation, S.-Y.G. and W.-S.K.; Writing—review and editing, S.-Y.G. and W.-S.K.; Visualization, S.-Y.G.; Supervision, Y.-J.K. and W.-S.K.; Project administration, Y.-J.K.; Funding acquisition, Y.-J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Korea Basic Science Institute (National research Facilities and Equipment Center) grant funded by the Ministry of Science and ICT (No. RS-2024-00404853).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Kim, J.-H.; Lee, C.-Y.; Cho, Y.-H.; Yu, Z.; Kim, K.-M.; Yang, Y.-J.; Nam, J.-S. Potato Farming in the United States and South Korea: Status Comparison of Cultivation Patterns and Agricultural Machinery Use. *J. Biosyst. Eng.* **2024**, *49*, 252–269. [[CrossRef](#)]
2. Raikwar, S.; Tewari, V.K. Development of Transmission Control Algorithm for Power Shuttle Transmission System for an Agricultural Tractor. *J. Biosyst. Eng.* **2022**, *48*, 136–151. [[CrossRef](#)]
3. Simikić, M.; Dedović, N.; Savin, L.; Tomić, M.; Ponjičan, O. Power Delivery Efficiency of a Wheeled Tractor at Oblique Drawbar Force. *Soil Tillage Res.* **2014**, *141*, 32–43. [[CrossRef](#)]
4. Ahn, D.-V.; Kim, K.; Choi, K.; Lee, J.W.; Kim, J.-G.; Yu, J.; Kim, H.-S.; Seo, J.; Park, Y.-J. Effect of Clutch Control to Improve Launch Quality for a Power Shuttle Tractor during Launching. *Comput. Electron. Agric.* **2024**, *224*, 109235. [[CrossRef](#)]
5. Kwon, D.; Ahn, D.-V.; Kim, J.-G.; Park, Y.-J. Effect Analysis of Motor Power Characteristics on the Energy Consumption of Dual Motor Driven Powertrain for Electric Tractor. *J. Biosyst. Eng.* **2024**, *49*, 465–475. [[CrossRef](#)]
6. Rajabi-Vandecchali, M.; Abbaspour-Fard, M.H.; Rohani, A. Development of a Prediction Model for Estimating Tractor Engine Torque Based on Soft Computing and Low Cost Sensors. *Measurement* **2018**, *121*, 83–95. [[CrossRef](#)]
7. Janulevičius, A.; Juostas, A.; Pupinis, G. Tractor’s Engine Performance and Emission Characteristics in the Process of Ploughing. *Energy Convers. Manag.* **2013**, *75*, 498–508. [[CrossRef](#)]
8. Zhu, S.; Wang, L.; Zhu, Z.; Mao, E.; Chen, Y.; Liu, Y.; Du, X. Measuring Method of Slip Ratio for Tractor Driving Wheels Based on Machine Vision. *Agriculture* **2022**, *12*, 292. [[CrossRef](#)]
9. Kara, S.; Karadirek, I.E.; Muhammetoglu, A.; Muhammetoglu, H. Real Time Monitoring and Control in Water Distribution Systems for Improving Operational Efficiency. *Desalination Water Treat.* **2016**, *57*, 11506–11519. [[CrossRef](#)]
10. Kim, J.; Lee, J.; Kim, D.; Choi, C.; Lee, M.; Kim, H.S. Developing a Prediction Model (Heavy Rain Damage Occurrence Probability) Based on Machine Learning. *J. Korean Soc. Hazard Mitig.* **2019**, *19*, 115–127. [[CrossRef](#)]
11. Kumari, A.; Reheman, H. Tillage Operation with a Tractor Drawn Rotavator Using an Embedded Advisory System for Minimizing Fuel Consumption. *J. Biosyst. Eng.* **2023**, *48*, 487–502. [[CrossRef](#)]
12. Oduma, O.; Oluka, S.I.; Eze, P.C. Effect of Soil Physical Properties on Performance of Agricultural Field Machinery in South Eastern Nigeria. *Agric. Eng. Int. CIGR J.* **2018**, *20*, 25–31.
13. Al-Shammary, A.A.G.; Caballero-Calvo, A.; Fernández-Gálvez, J. Evaluating the Performance of a Novel Digital Slippage System for Tractor Wheels Across Varied Tillage Methods and Soil Textures. *Agriculture* **2024**, *14*, 1597. [[CrossRef](#)]
14. Siddique, A.A.; Baek, S.-M.; Baek, S.-Y.; Jeon, H.-H.; Park, J.-D.; Park, M.-J.; Yang, C.-W.; Park, M.J.; Kim, Y.-S.; Kim, W.-S.; et al. Effect of Motor Speeds on Traction Performance of a Single-Motor Electric Tractor at Various Gear Stages during Plow Tillage. *Sci. Rep.* **2023**, *13*, 56590. [[CrossRef](#)]
15. Janulevičius, A.; Damanauskas, V. Prediction of Tractor Drive Tire Slippage under Different Inflation Pressures. *J. Terramech.* **2022**, *101*, 23–31. [[CrossRef](#)]
16. Ge, J.; Peng, S.; Cao, C.; Fang, L.; Qin, K. Numerical Prediction of Tractive Performance of Track–Soil Interaction System through Different Grouser Heights. *Eng. Agrícola* **2022**, *42*, 19–24.
17. Min, Y.-S.; Kim, Y.-S.; Lim, R.-G.; Kim, T.-J.; Kim, Y.-J.; Kim, W.-S. The Influence of Soil Physical Properties on the Load Factor for Agricultural Tractors in Different Paddy Fields. *Agriculture* **2023**, *13*, 2073. [[CrossRef](#)]
18. Alhassan, E.A.; Olaoye, J.O.; Lukman, A.F.; Adekanye, T.A.; Abioye, O.M. Statistical Modelling of a Tractor Tractive Performance during Ploughing Operation on a Tropical Alfisol. *Open Agric.* **2024**, *9*, 20220282. [[CrossRef](#)]
19. Al-Dosary, N.M.N.; Alnajjar, F.M.; Aboukarima, A.E.W.M. Estimation of Wheel Slip in 2WD Mode for an Agricultural Tractor during Plowing Operation Using an Artificial Neural Network. *Sci. Rep.* **2023**, *13*, 59975. [[CrossRef](#)]

20. Angelucci, L.; Varani, M.; Pinet, F.; Martin, V.; Vertua, A.; Molari, G.; Mattetti, M. The Role of Tyres and Soil Conditions in Enhancing the Efficiency of Agricultural Tractors. *Soil Tillage Res.* **2025**, *251*, 106570. [[CrossRef](#)]
21. Almaliki, S.A.; Himoud, M.S.; Muhsin, S.J. Mathematical Model for Evaluating Slippage of Tractor under Various Field Conditions. *Basrah J. Agric. Sci.* **2021**, *34*, 49–59. [[CrossRef](#)]
22. Seifu, Y.; Hiresamhir, S.; Tole, S.; Yilak, A. Evaluation of Tractor Field Performance Using Visual Basic Programming for Agricultural Farm Lands. *Am. J. Agric. Sci. Eng. Technol.* **2023**, *7*, 25–35. [[CrossRef](#)]
23. Al-Mastawi, K.E.; Dahham, G.A.; Yahya, L.M. Effects of Soil Moisture Content, Tire Inflation Pressure, and Tillage Speed on Tractive Performance of 2WD Tractor in Northern Iraq. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 409–418.
24. Kim, W.S.; Kim, Y.J.; Baek, S.Y.; Baek, S.M.; Kim, Y.S.; Kim, Y.K.; Choi, I.S. Traction Performance Evaluation of a 78-kW-Class Agricultural Tractor Using Cone Index Map in a Korean Paddy Field. *J. Terramech.* **2020**, *91*, 285–296. [[CrossRef](#)]
25. Shafaei, S.M.; Loghavi, M.; Kamgar, S. Feasibility of Implementation of Intelligent Simulation Configurations Based on Data Mining Methodologies for Prediction of Tractor Wheel Slip. *Inf. Process. Agric.* **2019**, *6*, 183–199. [[CrossRef](#)]
26. Najafi, G.; Ghobadian, B.; Tavakoli, T.; Buttsworth, D.R.; Yusaf, T.F.; Faizollahnejad, M. Performance and Exhaust Emissions of a Gasoline Engine with Ethanol Blended Gasoline Fuels Using Artificial Neural Network. *Appl. Energy* **2009**, *86*, 630–639. [[CrossRef](#)]
27. Al-Sager, S.M.; Almady, S.S.; Marey, S.A.; Al-Hamed, S.A.; Aboukarima, A.M. Prediction of Specific Fuel Consumption of a Tractor during the Tillage Process Using an Artificial Neural Network Method. *Agronomy* **2024**, *14*, 492. [[CrossRef](#)]
28. TYM Co. TYM Tractors Official Website. TYM Official Website. Available online: <https://tym.world> (accessed on 15 September 2025).
29. Spectrum Technologies Inc. FieldScout TDR 350 Soil Moisture Meter (Case Included). Spectrum Technologies Official Website. Available online: <https://www.specmeters.com/FieldScout-TDR350-Soil-Moisture-Meter> (accessed on 15 September 2025).
30. Spectrum Technologies Inc. FieldScout SC900 Soil Compaction Meter. Spectrum Technologies Official Website. Available online: <https://www.specmeters.com/FieldScout-SC900-Soil-Compaction-Meter> (accessed on 15 September 2025).
31. ASABE Standards EP542; Procedure for Using and Reporting Data Obtained with the Soil Cone Penetrometer. American Society of Agricultural and Biological Engineers: St. Joseph, MO, USA, 2019.
32. Sima, N.Q.; Harmel, R.D.; Fang, Q.X.; Ma, L.; Andales, A.A. A modified F-test for evaluating model performance by including both experimental and simulation uncertainties. *Environ. Model. Softw.* **2018**, *104*, 236–248. [[CrossRef](#)]
33. Tadić, V.; Radočaj, D.; Jurišić, M. Machine Learning Methods for Evaluation of Technical Factors of Spraying in Permanent Plantations. *Agronomy* **2024**, *14*, 1977. [[CrossRef](#)]
34. Dhanka, S.; Maini, S. A Hybridization of XGBoost Machine Learning Model by Optuna Hyperparameter Tuning Suite for Cardiovascular Disease Classification with Significant Effect of Outliers and Heterogeneous Training Datasets. *Int. J. Cardiol.* **2024**, *420*, 132757. [[CrossRef](#)]
35. Lundberg, S.M.; Lee, S. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.
36. Hancock, J.T.; Khoshgoftaar, T.M. CatBoost for big data: An interdisciplinary review. *J. Big Data* **2020**, *7*, 94. [[CrossRef](#)]
37. Ahn, J.M.; Kim, J.; Kim, K. Ensemble Machine Learning of Gradient Boosting (XGBoost, LightGBM, CatBoost) and Attention-Based CNN-LSTM for Harmful Algal Blooms Forecasting. *Toxins* **2023**, *15*, 608. [[CrossRef](#)]

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.