https://jksus.org

Journal of King Saud University – Science

Articles in Press



Research Article

A study on efficiency estimation of water based PV/T systems with machine learning methods

Merve Demirci[a,*], Rahim Aytug Ozer[b]

[a]Department of Electrical and Electronics engineering, Faculty of Engineering and Architecture, Kafkas Univeristy, Kars, Turkey
[b]Department of Aerospace Engineering, Faculty of Engineering, Architecture, Erzurum Technical University, Erzurum, Turkey

ARTICLE INFO

Keywords:
Cooling effect
Data pre-processing
Efficiency
Machine learning
Photovoltaic thermal systems

ABSTRACT

With the inevitable increase in energy consumption, there has been a growing interest in renewable energy sources. Therefore, efforts to increase the efficiency of solar panels have gained significant importance. In experimental studies conducted on Photovoltaic/thermal (PV/T) systems in the literature, the obtained data have been used in theoretical and mathematical calculations to derive efficiency expressions. However, deviations in the calculated efficiency values occur when environmental conditions and input parameters change. Therefore, in this study, efficiency prediction has been performed using machine learning (ML) methods to eliminate deviations and uncertainties. The input parameters, including mass flow rate, ambient temperature, inlet fluid temperature, cell absorptance, PV/T surface area, and solar radiation, along with a dataset collected from the literature, were used with ML methods such as Support vector regression (SVR), Linear regression (LR), Gaussian process regression, Stepwise LR, and artificial neural network (ANN) algorithms. To prepare the data for the algorithms, a preprocessing stage was applied, during which various data preprocessing methods, including min-max normalization, z-score normalization, and logarithmic transform + standardization, were implemented. According to the results, the highest prediction performance was achieved when min-max normalization was applied during the preprocessing stage, with the support vector regression algorithm yielding an root mean square error (RMSE) value of 0.225 and an R^2 value of 0.98018. Furthermore, according to the results of the F-test, it was concluded that the most influential parameter on efficiency is the PV/T surface area.

1. Introduction

The increase in the world population and technological advances cause energy consumption to constantly increase (Ozer *et al.*, 2023; Ozer *et al.*, 2025). The limited availability of fossil fuel reserves has made it necessary to turn to renewable energy sources in order to meet this increasing energy demand. Renewable energy, when compared to fossil fuels, offers a more environmentally friendly and less harmful alternative to human health, thus contributing significantly to environmental protection and sustainability while meeting energy needs. In this context, renewable energy sources are gaining importance as both an effective and efficient solution for achieving sustainable development. In this context, maximizing the electricity generation efficiency of PV panels stands out as a critical research topic that should be focused on.

The electrical energy produced by photovoltaic (PV) panels depends on many parameters related to the panel and the environment; the most important of these is the cell temperature of the panel (Joshi *et al.*, 2009; Hamada *et al.*, 2024; Swese *et al.*, 2022). An increase in cell temperature results in an increase in internal resistance and hence a decrease in efficiency (Swese *et al.*, 2022). New systems designed to remove waste heat from solar radiation have the potential to prevent the decline in electrical efficiency. For this purpose, PV/T systems have been developed. PV/T systems not only increase electrical efficiency by keeping the panel temperature within design limits, but also allow the cooling liquid to be heated and used for different purposes by absorbing waste heat. As a result of this dual benefit from solar panels, the total efficiency of the system increases between 4% and 12% (Ramos *et al.*, 2017). PV/T system designs are designed as open circuit (Yang and Athienitis, 2014) and closed-circuit systems that operate by recirculating water (Yu *et al.*, 2019).

Experimental and numerical studies on the variation of thermal and electrical efficiency of PV/T systems depending on variable input parameters are available in the literature. In these studies, the relationship between thermal, electrical and total efficiency and variable parameters was tried to be estimated (Salman *et al.*, 2022; Ma *et al.*, 2021; Yao *et al.*, 2020; Bait, 2019; Panagiotidou *et al.*, 2020). Solar radiation (Azad *et al.*,2024; Gharaee *et al.*,2024), PV panel angle, coolant type (Azad *et al.*,2024; Sopian *et al.*,2020; Sangeetha *et al.*, 2021; Melaibari *et al.*, 2024; Sheikholeslami and Khalili, 2024), inlet temperature-speed (Sheikholeslami and Khalili, 2024), fluid flow rate (Sopian *et al.*,2020; Melaibari *et al.*, 2024; Muslim *et al.*, 2020; Basuhaib *et al.*, 2024), wind speed (Gharaee *et al.*,2024), heat sink geometrical features (Muslim *et al.*, 2020; Hosseinzadeh *et al.*, 2018; Kiwan and Khlefat, 2021), latitude-longitude where the PV panel is positioned and ambient temperature are the prominent parameters

***Corresponding author:**
E-mail address: merve.demirci@kafkas.edu.tr (M Demirci)

Received: 15 December, 2025 **Accepted:** 06 March, 2026 **Epub Ahead of Print:** 29 June, 2026 **Published:** ***

DOI: 10.25259/JKSUS_1980_2025

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, transform, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

© 2026 Journal of King Saud University – Science - Published by Scientific Scholar

affecting the PV/T system efficiency. The performance analysis of flat plate PV panels is based on the model known as the Hottel-Whillier model, which was adapted to the PV/T combined system analysis by Florschuetz (Farahani *et al.,* 2025). Many studies consist of evaluations based on data obtained from theoretical or experimental studies. For example, the performance analysis of the PV/T system designed with a tube model containing nanofluid was numerically conducted by Sheikholeslami and Khalili (2024). In this study, heat pipes were used and the fluid inlet velocity was changed between 0.01 and 0.18 m/s. The fluid inlet temperature was kept constant as 30°C and the solar radiation value as 800 W/m^2. The maximum exergy was determined as 15.32%, electrical efficiency as 14.05% and thermal efficiency as 55.22%. In the study conducted by Muslim *et al.* (2020) in Iraq, the effects of panel orientation and reverse flow on the performance of a PV/T system using a closed water cycle under climatic conditions were investigated. It was stated that maximum electrical efficiency was improved by 11.5%, maximum thermal efficiency by 58.64% and total efficiency by 66.87%. Hosseinzadeh *et al.* (2018) investigated the effect of Phase change materials (PCM) on system performance by performing energy and exergy analysis of PV thermal systems. Kiwand and Khlefat (2021), investigated the variation in PV panel performance with the use of porous material. According to the results obtained, the maximum efficiency was determined as 9.19% in the case of using five porous structures, the minimum efficiency was determined as 6.73% in the case of using three porous structures and 8.34% in the case of using porous layers.

In literature studies, theoretical equations obtained from experimental or mathematical models are used to predict system performance (Mojumder *et al.,* 2017). Theoretical equations are based on relating the effects of independent variables on the dependent variable with correction coefficients, taking into account the experimental parameters of a certain physical model, and a pool of data sets. Therefore, when the environmental conditions and parameters change, the system does not conform to the theoretical equations. To overcome this issue, Machine learning (ML) regression methods have been utilized in recent years.

One of the studies conducted in this field was carried out by Farahani *et al.,* (2025). In this study, a combination of jets and foamed gradients was used to increase the electrical efficiency of air-based PV/T systems, and subsequently, ML methods such as gauss process regression (GPR) and adaptiv neuro-fuzzy inference system (ANFIS) were employed to predict electrical efficiency using the obtained data. The best predictions were obtained using the GPR method. In another study, Zamen *et al.,* (2019) used least square support vector machine (LS-SVM), multi layer perceptron neural network (MLPNN), and ANFIS to predict the thermal efficiency of PV/T systems based on the input temperature, circulation flow rate, and solar radiation parameters. Jackar *et al.,* (2023) used parameters such as Reynolds number, time and fluid concentration in the multi layer perceptron (MLP) algorithm to predict the electrical and thermal performance of PV/T systems containing water-based nanofluids. They also compared the performance of MLP with LR, Decision Tree, and SVM, which are frequently used in the literature. Diwania *et al.,* (2024), analyzed the thermo-electric performance of PVT systems operating with deionized water and Fe-H_2O nanofluids depending on the real input data of solar radiation and ambient temperature and the day-ahead forecast data obtained by the GPR method. Shahsavar *et al.,* (2020) evaluated the system performance in terms of exergy using the length, width, depth and mass flow rate of the channel under the photovoltaic (PV) modules in building integrated PV/T systems. Multiple LR, MLP, lazy. IBK, Random Forest, Radial Basis Function Regressor, SVM and Random tree algorithms were used and the performance of the algorithms was compared with the performance evaluation criteria. Ahmadi *et al.,* (2020) used Neural Network, Least Square SVM and Neuro-Fuzzy methods in their study to estimate the thermal performance of the PV/T system using the inlet temperature, fluid flow rate, solar radiation and solar heat parameters. Caner *et al.,* (2011) designed two types of solar air collectors in their study. They utilized a Levenberg-Marquardt-based MLP to predict the thermal performance of the system using the obtained experimental data. The predicted thermal performance values were compared with the actual values. Examination of existing studies in the literature reveals that they primarily focus on thermal efficiency. Furthermore, the datasets used in thermal efficiency estimation contain a small number of input data. In this study, six different input parameters were used to estimate electrical efficiency using dataset-based PV/T modeling, a topic that has been the focus of attention in the literature in recent years. In addition, preprocessing was applied to the data to improve estimation accuracy, and comparative results were obtained using multiple algorithms, resulting in a reliable and robust model.

The key contributions of this study are following:

- In PV/T systems, ML regression methods are used to estimate electrical efficiency depending on the parameters.
- For electrical efficiency estimation, mass flow rate, ambient temperature, fluid temperature, PV cell absorbance, PV surface area and solar irradiation data are collected from the literature from the parameters affecting electrical efficiency in water-based PV/T systems.
- In order to prepare the data set for the algorithms, different preprocessing methods are used to scale the data differently. For this, min-max normalization, Z-Score normalization, logarithmic transformation + standardization methods that are available in the literature and widely used are used for data preprocessing.
- F-test is applied to determine how effective the features in the data set are in electrical efficiency estimation.
- Electrical efficiency estimation is performed using LR, Support vector regression (SVR), Artificial neural network (ANN), GPR and Linear stepwise regression (LSR) from ML regression methods. The performances of the methods are obtained as RMSE, R-squared (R^2), Mean squared error (MSE) and Mean absolute error (MAE) performance metrics. These processes are repeated and compared with the raw data set and data sets that have undergone different preprocessing.
- In order to prevent overfitting in the estimation process, electrical efficiency is estimated by applying cross validation (cv=5).
- When the obtained results are compared, the SVR algorithm has the best performance with RMSE =0.225 R^2 = 0.98018 values with the data applied to min-max normalization. The performance of the applied method is shown when compared with the studies in the literature.

2. Materials and Methods

2.1 PV/T system

In this study, five ML models were used to predict the electrical efficiency of PV/T systems. The system performance was analyzed based on the laws of thermodynamics. Therefore, the governing equations utilized are presented in this section.

PV/T systems are combined systems used to obtain both electrical and thermal energy from solar energy. To determine the amount of energy that a PV panel can produce electrically, it is important to define the relationship between electrical current and voltage. Fig. 1 illustrates the relationship between voltage and current in a PV cell. The maximum voltage that can be generated from the PV panel using the Carnot machine approach operating between the solar and cell temperature range is given in Eq. (1) (Saloux *et al.,* 2013).

$$V_{o,max} = \left(1 - \frac{T_{PV}}{T_{sun}}\right) \frac{E_{gap}}{q_e} \tag{1}$$

Here, T_{PV} is the PV panel cell temperature, q_e is the electronic electric charge, and E_{gap} is the semiconductor bandgap energy. This equation is constrained for specific temperature conditions, and variations in these parameters non-linearly affect the resultant value.

Energy balance equations are used when performing numerical analysis in PV systems. The PV cell energy conservation equation is given in Eq. (2). In this equation, ε_c represents the absorptivity coefficient of the PV cell, σ_c is the cell transmissivity coefficient, β_c denotes the packing factor coefficient, and η_c indicates the instantaneous collector

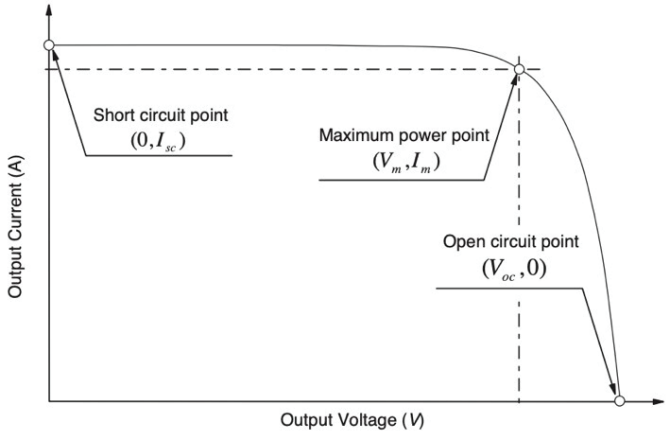


Fig. 1. V-I variation for PV cell.

efficiency. In this equation, $G(t)$, represents the solar irradiation intensity, $U_{t\,c,a}$, denotes the total heat transfer coefficient from the PV panel to the environment, $h_{c,p}$, is the heat transfer coefficient from the PV cell to the panel, H represents the panel width, and T_c, T_a and T_p denote the PV cell temperature, ambient temperature, and plate temperature, respectively (Gharaee et al., 2024).

$$\varepsilon_c\sigma_c\beta_c G(t)Wdx = \left[U_{t\,c,a}\left(T_c - T_a\right) + h_{c,p}\left(T_c - T_p\right)\right]Hdx + \eta_c\sigma_c\beta_c G(t)Hdx \tag{2}$$

The energy balance equation for the absorber surface under the PV module is given in Eq. (3). Here, ε_p represents the absorptivity coefficient of the surface, σ_g is the transmissivity coefficient of the glass surface, T_f denotes the fluid temperature, and $h_{p,f}$ is defined as the heat transfer coefficient from the absorber surface to the cooling fluid.

$$\varepsilon_p\sigma_g^2\left(1-\beta_c\right)G(t) + h_{c,p}\left(T_c - T_p\right) = h_{p,f}\left(T_p - T_f\right) \tag{3}$$

The energy balance equation for the fluid circulating in the heat sink tubes can be written as in Eq. (4). In this equation, $\dot{m}_f$ represents the fluid flow rate, C_f denotes the fluid heat capacity, and K' indicates the efficiency coefficient of the collector for a flat plate.

$$\dot{m}_f C_f \frac{dT_f}{dx}dx = K'h_{p,f}\left(T_p - T_f\right)Hdx \tag{4}$$

The energy conservation equation in a PV/T system is an important step in determining the system efficiency. In these systems, the total input energy is solar radiation and is equal to the electrical energy produced in the PV panel, the heat losses in the connection elements and the thermal energy absorbed by the heat sink and/or cooling liquid, depending on the design type Eq. (5). The total, electrical and thermal efficiency of the PV/T system can be calculated with the help of Eq. (6-8).

$$E_i = E_e + E_{th} + E_{loss} \tag{5}$$

$$\eta_{tot} = \frac{E_e + E_{th}}{E_i} \tag{6}$$

$$\eta_{th} = \frac{E_{th}}{E_i} \tag{7}$$

$$\eta_e = \frac{E_e}{E_i} \tag{8}$$

To obtain the efficiency equations in PV/T systems, it is necessary to calculate the thermal equations. These systems, which involve different heat transfer mechanisms, represent a complex engineering problem. Therefore, accurately calculating performance in these systems is challenging. In addition, as can be seen from the equations used to analyze PVT systems based on the laws of thermodynamics, the input parameters are the measured quantities. For this reason, significant deviations can be observed due to uncertainties resulting from reading and measurement devices. Due to these deviations, an unacceptable difference occurs between the actual and theoretical efficiencies of PV/T systems. To overcome all these problems, ML methods are useful and have become an indispensable approach in recent years due to their ability to make accurate predictions. In the following section, the ML methods used in the study are briefly described.

2.2 ML regression algorithms

2.2.1 LSR

The aim of the LSR method is to optimally combine the independent variables when making predictions of the dependent variable. For this purpose, it uses forward addition and backward deletion methods together. As a result of successive operations, it decides whether to include or remove variables in the estimation process as a result of statistical tests (Chen *et al.,* 2013; Ma *et al.,* 2016).

The stepwise procedure begins by gradually adding variables to the previous model and then tests the suitability of the variables to the model using statistical tests. The operations are repeated iteratively in order to obtain meaningful results with the new variable added. With each added variable, the predictive capability of the model may change. In cases where the added variable is unimportant, the variable is deleted using the backward elimination method. These processes continue until no variables are added or removed. The stepwise procedure and statistical tests conclude when the final model is obtained (Chen *et al.,* 2013; Ma *et al.,* 2016; Shi *et al.,* 2023).

2.2.2 GPR

GPR is a nonparametric probabilistic kernel estimation algorithm that has been widely used in recent years. The GPR method is not only used for prediction but also provides a confidence interval to quantify the uncertainty of the prediction at each point (Sharifzadeh *et al.,* 2019; Zhao *et al.,* 2011).

The GPR algorithm, which applies the Gaussian Process while performing the estimation process, calculates the mean and covariance vector for each step it applies. In this method, the initial GPR is first created to reveal the unknown relationship in the training dataset, then the limited initial GPR is updated using the training samples and the final model GPR is obtained. To learn the GPR model, it is necessary to know the kernel parameters (Zhao *et al.,* 2011; Williams and Rasmussen, 2006). This algorithmic approach is widely used in non-linear and complex regression and classification problems (Hong and Zhou, 2012).

2.2.3 SVR

SVR is a ML methodology that operates according to learning theory, introduced by Vapnik and Cortes in 1995 (Cortes *et al.,* 1995). It maps nonlinear low-dimensional data to high-dimensional space and presents a linear model to classify this data. This is the main purpose of SVR (Gharaee *et al.,*2024).

Support vector machines (SVMs) used in classification processes define a hyperplane that aims to separate classes and maximize the margin between them. In SVR, the aim is to minimize the margin between the data on the right and left sides of this hyperplane and to collectively reduce the deviation of all sample points from this hyperplane. In SVR, the model is managed and adapted to the problem using hyperparameters such as penalty parameters and kernel

functions. One of the advantages of the SVR algorithm is that the computational complexity of the model is independent of the number of input parameters. In addition, SVR was chosen due to its accuracy and excellent generalizability (Gharaee *et al.*,2024; Huang et al., 2023).

2.2.4 ANN

ANNs are widely used for regression, classification, correlation, and statistical analysis. ANN is more capable than humans and other computational techniques in deriving conclusions from complex data and generating novel solutions. It can extract information from data that may be incomprehensible to any expert, and it can act as a specialist in that field using this information.

ANN structures consist of input, hidden, and output layers with varying numbers of neurons. The input layer takes in the dataset, it is processed in the hidden layer, and the results are presented to the user through the output layer (Yilmaz and Kaynar, 2011; Zain *et al.*, 2012; Demirci *et al.*, 2022). In the MLP model, data received through the input layer is processed in one or more hidden layers using an activation function and then transferred to the subsequent output layer. Each layer performs a series of operations, and results are obtained from the output layer through weighted connections (Yilmaz and Kaynar, 2011).

2.2.5 LR

In statistical problems, one of the methods used to predict the relationship between several different variables is the LR algorithm. The goal is to reveal the relationship between several independent variables and a dependent variable to make the most accurate prediction. Single variable LR makes predictions based on a single variable, while multi variable regression attempts to make predictions using two or more variables. Here, a linear fit is attempted by keeping all variables constant except for one, in order to define the relationship between them (Chakraborty and Goswam, 2017; Tatachar, 2021).

2.3 Performance evaluation metrics

In regression problems, various measurement metrics are utilized to evaluate the performance of algorithms. These performance evaluation metrics include: Root Mean Square Error (RMSE), (R^2), MAE, and MSE (Tatachar, 2021; Shi *et al.*, 2023).

2.3.1 MSE

This performance metric is defined as the average of the squares of the differences between the actual and predicted values in a dataset. MSE always takes on positive values, and the closer it is to 0, the better the quality of the predictions. The equation for MSE is provided in Eq. (9). Here, n represents the number of predictions, y_i denotes the actual values and $\hat{y}_i$ indicates the predicted values.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{9}$$

2.3.2 RMSE

It is the square root of the average of the squared errors from all predictions. It indicates how closely the data points fit the best-fitting line. The RMSE formula provided in Eq. (10) compares the actual values with the predicted values. Here, n represents the number of predictions, y_i denotes the actual values, $\hat{y}_i$ indicates the predicted values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{10}$$

2.3.3 MAE

It is the average of the absolute differences between the actual values and the predicted values in a dataset. The formula for MAE is provided in Eq. (11). Here, n represents the number of predictions, y_i denotes the actual values, $\hat{y}_i$ indicates the predicted values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{11}$$

2.3.4 R^2

In a regression problem, R^2 indicates the goodness of fit of the model, measuring the extent to which the actual values overlap with the predicted values. The R^2 value ranges from 0 to 1, with a value closer to 1 indicating a better fit of the model. It provides information for model selection by allowing the comparison of different regression algorithms in terms of their fit. The equation for the R^2 evaluation metric is provided in Eq. (12). Here, SSR is Sum of Squares of residuals, TSS is Total Sum of Squares.

$$R^2 = 1 - \frac{SSR}{TSS} \tag{12}$$

2.4 Data pre-processing

In classification and regression problems related to ML methods, the characteristics of the dataset play a critical role in determining algorithm performance. Therefore, preprocessing the data is an important step to ensure that the data is appropriately prepared for classification and regression algorithms. In this study, the dataset was processed using min-max normalization, Z-score standardization, and logarithmic transformation + standardization methods. These preprocessing methods were chosen because they are widely used in data-driven energy system problems and modeling, and are effective in reducing numerical dominance among input parameters and mitigating the negative effects arising from variables with different scales. The effects of each preprocessing method on the performance of regression algorithms were obtained, the results were compared, and it was observed that different preprocessing strategies had a significant effect on model performance.

2.4.1 Min-max normalization

Min-max normalization can be performed using Eq. (13). In this equation; x represents the raw data, x_{min} denotes the minimum value of x, x_{max} indicates the maximum value of x. X_{norm}, represents the normalized data (Henderi *et al.*, 2021).

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{13}$$

2.4.2 Z-score normalization

Eq. (14) contains the formula for the standardization process. In this equation; x represents the raw data, μ denotes the mean value, σ indicates the standard deviation. X_{std}, represents the standardized data (Henderi *et al.*, 2021).

$$x_{std} = \frac{x - \mu}{\sigma} \tag{14}$$

2.4.3 Logarithmic transform+ standardization

Eq. (15) contains the formula for logarithmic transformation, while Eq. (16) presents the equation for the standardization process. X represents the raw data, x_{log} denotes the data that has undergone logarithmic transformation, and x_{std} represents the standardized data. Additionally, μ indicates the mean value, and, σ represents the standard deviation (Demirci *et al.*, 2023).

$$x_{log} = \log_{10} x \tag{15}$$

Table 1.
The characteristics of the data set.

	Mass flow rate	Solar radiation	Ambient temperature	Inlet Fluid temperature	PVT surface area	Cell absorptance	Electrical efficiency
Count	151.000	151.000	151.000	151.000	151.000	151.000	151.000
Mean	0.046	776.159	27.746	27.7457	0.784	0.8646	10.464
Std	0.036	141.401	3.555	3.555	0.408	0.032	1.601
Min	0.003	400.000	23.850	23.850	0.346	0.850	6.196
Max	0.170	1000.000	32.700	32.700	1.326	0.950	16.054

$$x_{std} = \frac{x_{log} - \mu}{\sigma} \tag{16}$$

2.5 Dataset

There are many parameters that affect system performance in PVT systems. These are; PV cell absorptivity coefficient, permeability coefficient, cell temperature, ambient temperature, fluid temperature, fluid flow rate, emissivity, mass flow rate, solar radiation and surface area. When the studies in the literature were examined, it was observed that there were six parameters that most affected the system performance. These parameters are: solar irradiation, mass flow rate, ambient temperature, fluid temperature, cell absorbance and surface area.

Solar irradiance (G) is one of the most critical parameters affecting the performance of PV/T systems, as it constitutes the primary energy source driving electricity generation in PV cells. In addition, the temperature difference between the PV/T system and the ambient environment plays a significant role in system performance; therefore, ambient temperature and inlet fluid temperature are also key influencing factors. The mass flow rate of the working fluid determines how effectively the waste heat generated within the PV/T system is absorbed, directly affecting the thermal behavior of the fluid. Furthermore, the absorptance of the PV cells governs the fraction of incident solar energy that can be converted, while the collector surface area defines the extent of solar radiation received and thus directly influences overall system performance. Building on the physical significance of these parameters, the governing equations presented above provide a comprehensive thermodynamic framework describing PV/T system behavior. In the context of this study, these equations primarily serve to establish the physical basis for selecting the input and output variables used in the ML models. Accordingly, physically meaningful parameters such as G, ambient temperature, inlet fluid temperature, mass flow rate, PV cell absorptance, and collector surface area were employed as input features, while electrical efficiency was defined as the output variable to be predicted. Through the ML approach, these inputs were supplied to the models, enabling a quantitative identification of the parameters that exert a more dominant influence on system efficiency. In this way, ML methods capture the complex and nonlinear relationships between thermodynamically related variables and system performance, thereby overcoming the limitations of purely analytical approaches under varying operating conditions.

In this study, a total of 151 data were collected from the literature regarding six input parameters affecting the performance of water-based PV/T systems. The dataset consists of 50 data from Abdul-Ganiyu *et al.,* (2021), 16 from Al-Shamani *et al.,* (2016), 36 from Abdullah *et al.,* (2020), 12 from Fudholi *et al.,* (2019), 5 from Ahmadinejad *et al.,* (2023) and 32 from Yeşildal *et al.,* (2022). Using these data, electrical efficiency estimates for PV/T systems were made. In the collected data set, the mass flow rate varies between 0.0027-0.138 kg/s, while the solar radiation varies between 500-1000 W/m^2. Ambient and fluid temperatures vary between 23.85°C-32.7°C. Surface area is between 0.3465-1.326 m^2. Cell absorbance parameter varies between 0.85-0.95. The characteristics of the content of the data set are presented in Table 1.

Fig. 2(a) presents the results of the F-test applied to the dataset. The F-test shows that PV/T Surface Area, Mass Flow Rate, and Cell Absorption parameters are the most statistically significant parameters in predicting electrical efficiency. Ambient Temperature, Inlet Fluid Temperature, and Solar Radiation parameters are less influential.

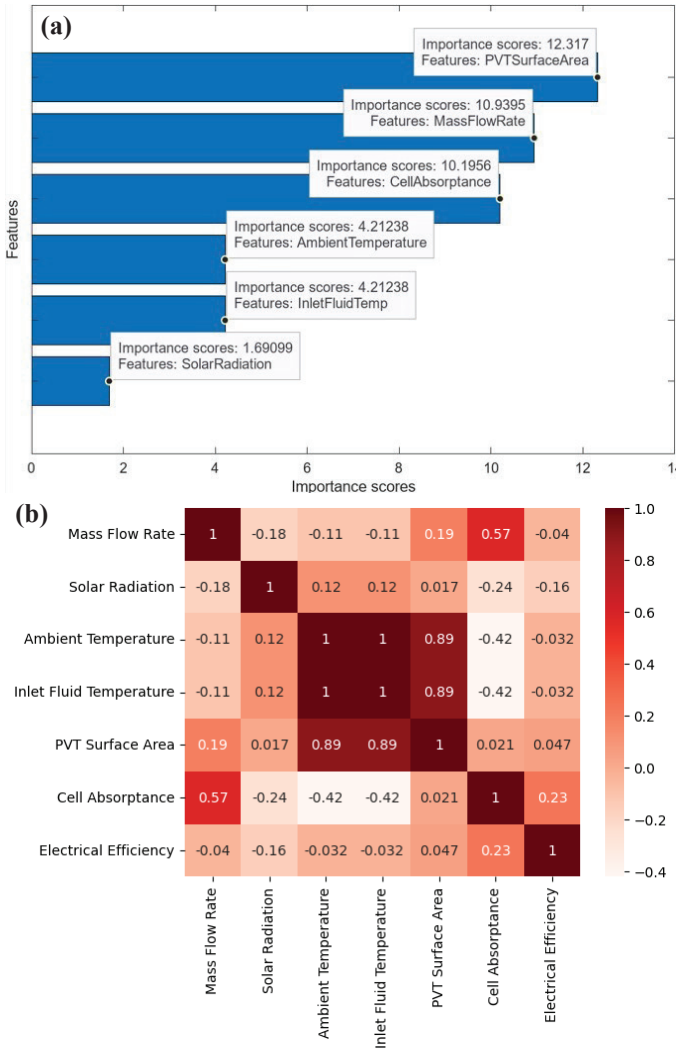


Fig. 2. (a) F test Scores. (b) Heatmap of correlation coefficients.

Fig. 2(b) presents the Heatmap generated based on the correlation between the input data. The correlation Heatmap reflects only linear relationships. As shown in Fig. 2(b), there is a low correlation between individual input parameters and electrical efficiency. This indicates that PVT system behavior is related to multivariate and nonlinear relationships, not individual and simple relationships.

The process steps applied in this study are shown in the flow chart given in Fig 3. First, data were collected from the literature and these data were subjected to a preprocessing step in order to prepare them for regression algorithms. Here, the data were processed with different preprocessing methods. Then, the ML regression algorithms to be used were selected and trained and tested with the processed data. Here, cross validation (cv = 5) was used. While the model was trained with a part of the dataset, the remaining part was tested, and this process continued until the entire dataset was tested. As a result of the training and testing process of the regression algorithms, prediction data were

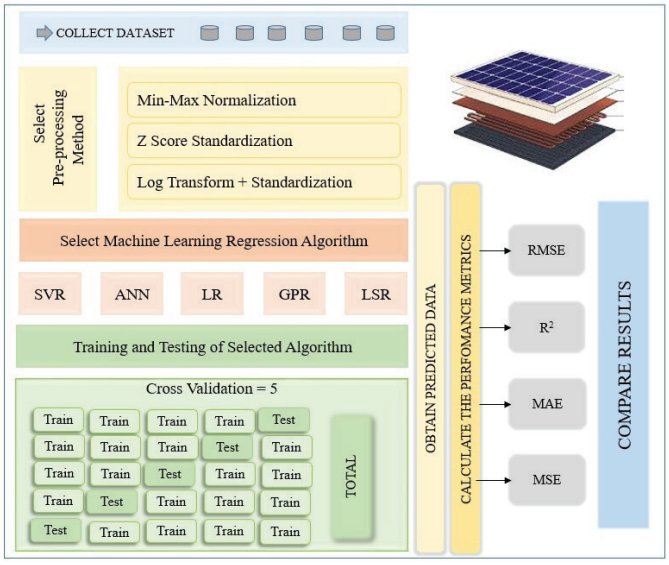


Fig. 3. Flow chart of study.

Table 2.
Data set scales for different preprocessing techniques.

	Mass flow rate	Solar radiation	Ambient tempera-ture	Inlet fluid tempera-ture	PVT surface area	Cell absorptance
Raw dataset	0.0027–0.138	500–1000	23.85–32.7	23.85–32.7	0.346–1.326	0.85–0.95
Min–Max	0–1	0–1	0–1	0–1	0–1	0–1
Z-score	−0.92–3.38	−2.66–1.58	−1.1–1.39	−1.1–1.39	−1.07–1.32	−0.45–2.63
Log T + Std	−3.3–2.06	−3.16–1.34	−1.15–1.39	−1.15–1.39	−1.14–1.24	−0.45–2.62

obtained, and in order to compare the performance of the algorithms, the performance evaluation metrics RMSE, R-Square, MSE and MAE parameters were calculated. Finally, the performance of the regression models was compared using the obtained performance metrics.

3. Results and Discussion

This section presents the results of analyses conducted to predict the electrical efficiency of water-based PV/T systems in four different scenarios. To ensure compatibility with ML algorithms, the dataset obtained from the literature was processed using three different preprocessing techniques in addition to the raw data, and a dataset was created for each case. The scaling characteristics of these datasets are summarized in Table 2. Regression algorithms were trained and tested for each case, and predicted efficiency values were obtained. Five-fold cross-validation was used to reduce overfitting and increase model robustness. The performance of the regression algorithms was evaluated and compared using multiple statistical measures; the results are presented both in tabular and graphical form. In this study, hyperparameter tuning for ML models was determined automatically

Table 3.
Hyperparameters of regression algorithms.

Algorithms	Hyperparameters
LR	Interactions linear Terms: Interactions Robust option: Off
LSR	Stepwise Linear Initial terms: Linear Upper bound on terms: Interactions Max. number of steps: 1000
SVR	Cubic SVM Kernel function: Cubic Kernel scale: Automatic Box constraint: Authomatic Epsilon: Auto
GPR	5/2 GPR Basis function: Constant Kernel function: Matern 5/2 Use isotropic kernel: Yes Kernel scale: Authomatic Signal standard deviation: Authomatic Sigma: Authomatic
ANN	Trilayered Neural Network Number of fully connected layers: 3 First layer size:10 Second layer size:10 Third layer size: 10 Activation: ReLU Iteration limit:1000 Regularization strength (lambda): 0

using the built-in optimization and analysis functions of the toolbox. The toolbox internally evaluates and analyzes multiple hyperparameter combinations within predefined and algorithm-specific ranges. Based on the analyses performed for multiple hyperparameters, it identifies the model with the best performance under a consistent validation framework. Consequently, even without the use of explicit and user-adjustable hyperparameter optimization algorithms, this process was automatically performed through iterative evaluations. This automated process reduced user bias and excessive computational complexity, ensuring objective hyperparameter selection. The parameters of the regression algorithms considered in this study are given in Table 3.

3.1 Case-1- raw dataset

In this case, electrical efficiency estimation was made with the selected regression algorithms LR, SVR, LSR, GPR and ANN using the raw dataset collected from the literature. Table 4 shows the performance evaluation metrics of the algorithms for the original dataset used in case-1, the algorithm's estimation speed (obs/sec) and training time (sec).

When the results obtained are compared, it is seen that the GPR model has the lowest RMSE value of 0.34778 when raw data is used, while the LSR model has the highest RMSE value of 0.4511. The R-Square value shows the ability of the model used to adapt to the problem and how much the estimated value coincides with the real value. The R^2 value closest to 1 is in the GPR model with 0.95354. This value shows that the most appropriate regression model for electrical efficiency estimation with the raw data set is GPR. It is seen that the GPR algorithm has a training time of 4.2706 sec and a prediction speed of 8100 obs/sec.

Table 4.
Comparison of different ML regression algorithm performance for case-1.

Regression algorithm	Data- processing	RMSE	MSE	R-square (R^2)	MAE	Prediction speed	Training time
LR	-	0.4242	0.1799	0,9309	0.2780	6800	0.6724
LSR	-	0.4511	0.2035	0.92184	0.30745	5900	6.5728
SVR	-	0.41672	0.17365	0.9333	0.205	4500	21.085
GPR	-	0.34778	0.120965	0.95354	0.1864	8100	4.2706
ANN	-	0.36561	0.13367	0.94866	0.21557	4600	8.676

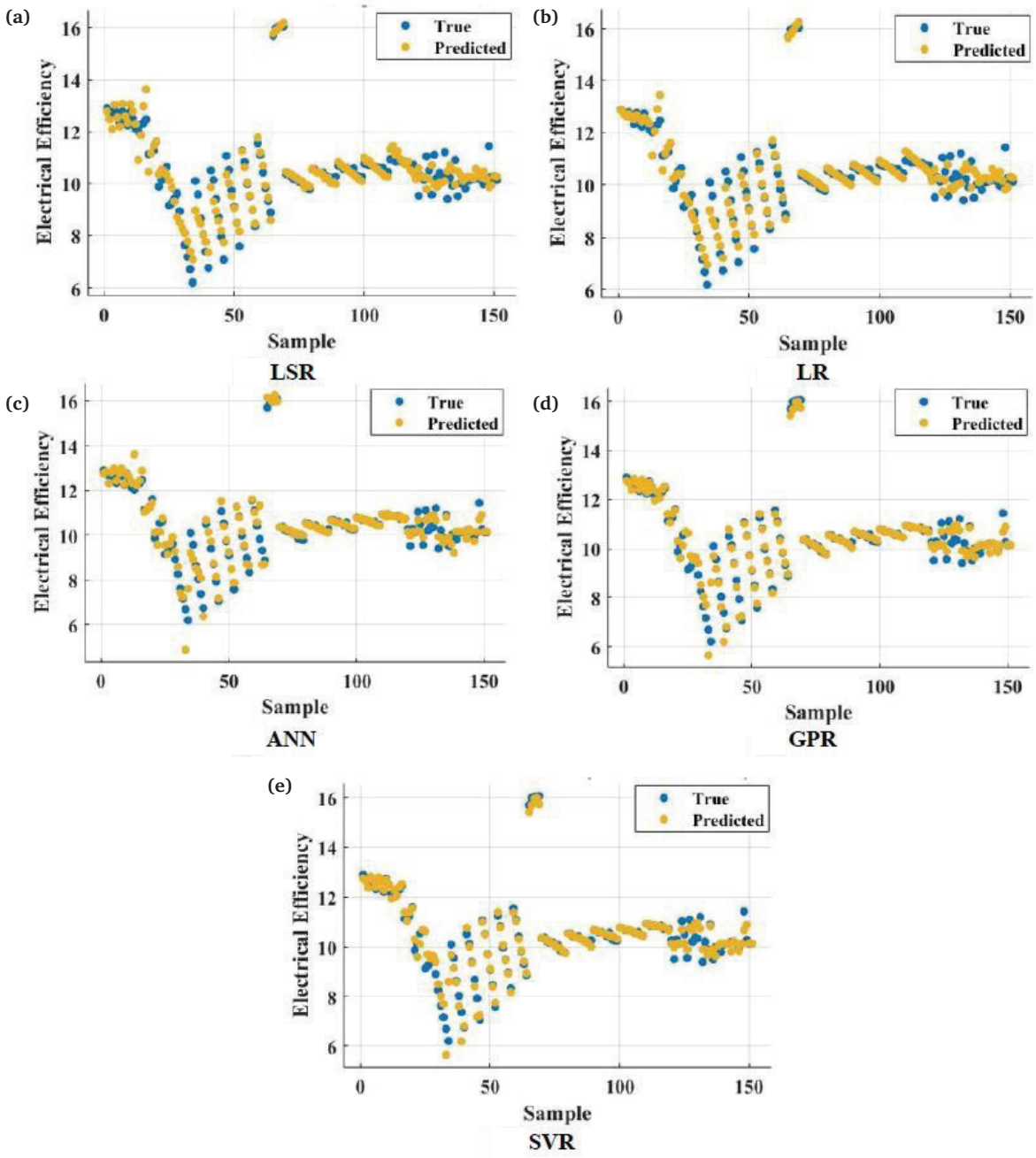


Fig. 4. Scatter plot of ML algorithms in case-1. (a) LSR, (b) LR, (c) ANN, (d) GPR, (e) SVR.

Fig. 4 shows graphs containing real and estimated data of regression algorithms. In these graphs, the distance between the electrical efficiency values estimated by the algorithms and the real values is given as a scatter plot in the data set used.

Fig. 5 presents predicted-true response graphs of regression algorithms. The better the performance of the regression algorithm, the closer the observation values are to the best prediction line. In the LR, LSR, ANN and SVR algorithms, it is seen that the predicted values are scattered around the true value line in a wider area and move away from the true value line. In contrast, it is seen that the predictions made using the GPR algorithm are positioned closer to the true value. This shows that the GPR algorithm is the most accurate algorithm in the prediction process made with the raw data set.

The electrical efficiency values estimated for each input parameter in the GPR algorithm with the highest prediction performance are presented in Fig. 6. These graphs graphically present the electrical efficiency value for each input parameter value, showing the range in which the input values change and their effect on the electrical efficiency. The predicted and real yield values in the graphs and the deviation between these values are presented as an error scatter plot indicated by the red line.

3.2 Case-2- min-max normalization

In this case, min-max Normalization process was applied as a data preprocessing method in order to prepare the data set used in the study for ML algorithms. Min-max normalization rescales all input variables to a fixed range [0,1], ensuring that each parameter contributes proportionally to the model generated during training. The aim here is to prepare the data for algorithms through scaling. The performance evaluation metrics, prediction speed, and training time values of the 5 ML regression algorithms used are presented in Table 5.

When the obtained results are compared, it is seen that the SVR algorithm has the lowest RMSE value, and the R^2 value, which shows

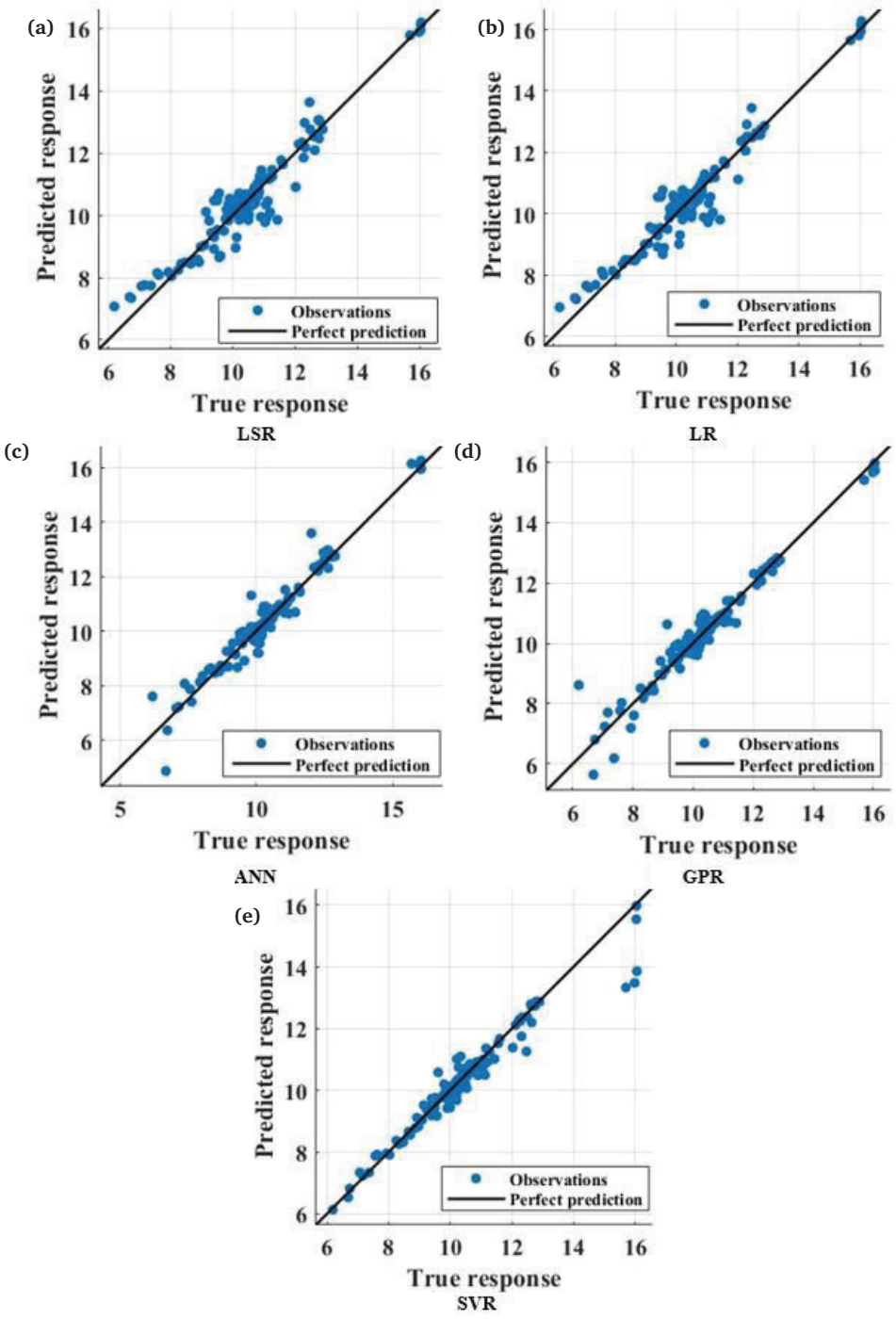


Fig. 5. Predicted and true response Scatter plot of ML algorithms in case-1. (a) LSR, (b) LR, (c) ANN, (d) GPR, (e) SVR.

the fit of the algorithm to the problem, is the highest with 0.98018. The RMSE value, which is 0.41672 in the raw data set, decreased to 0.225 when the data was preprocessed. The worst performance for the processed data set belongs to the LSR algorithm with a RMSE value of 0.44072. In addition, the fact that the lowest R^2 value is in LSR shows that the algorithm has the lowest performance and that the algorithm is the least fit to the problem.

Fig. 7 illustrates the scatter plots of the actual versus predicted values obtained from regression analyses performed using the dataset preprocessed with Min-Max normalization. The primary objective is to achieve a close match between the predicted and actual values, indicating high model accuracy. Among the algorithms evaluated, the SVR algorithm exhibits the highest level of agreement, as evidenced by the strong overlap between the predicted and actual data points in its scatter plot.

In Fig. 8, predicted-true response graphs of regression algorithms belonging to case-2 are presented. In LR, LSR, ANN and GPR algorithms, it is seen that the predicted values are scattered in a wider region around the true value line and even in some cases they are quite far from the true value line. In contrast, it is seen that the predictions made using the SVR algorithm are positioned in a way that they will be closest to the true value. This shows that the SVR algorithm is the most accurate algorithm in the prediction process made with the data subjected to the min-max normalization process.

The electrical efficiency values estimated for each input parameter in the SVR algorithm with the highest estimation performance are

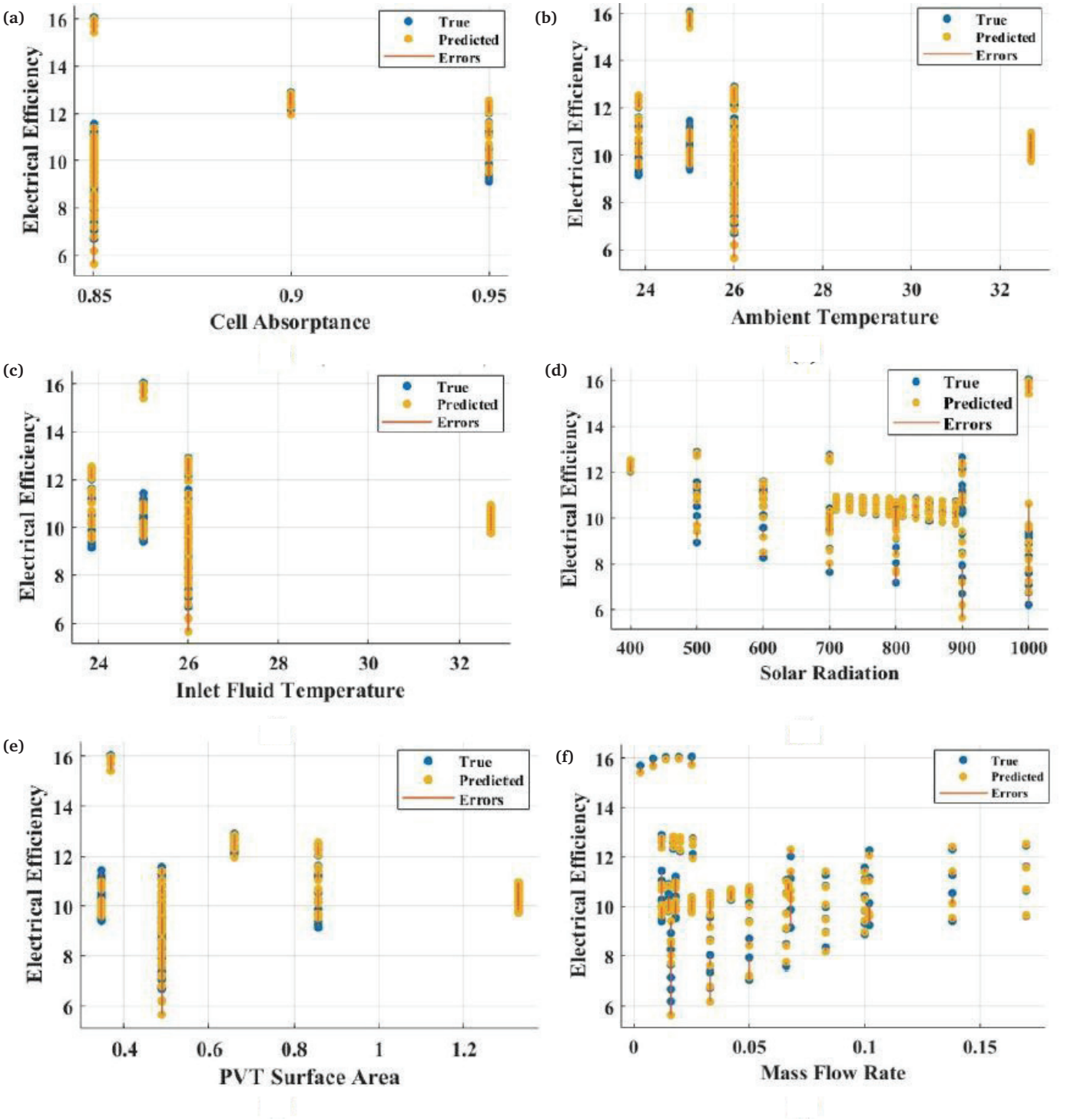


Fig. 6. True- predicted and errors values scatter graphics of different input variables of GPR in case-1. (a) Cell absorptance, (b) Ambient temperature, (c) Inlet fluid temperature, (d) Solar radiation, (e) PVT Surface Area, (f) Mass flow rate.

presented in Fig. 9. Here, the predicted scatter plot is given with the real data and the predicted data depending on the input parameters. What is important here is that the error value, which represents the difference between the real value and the graph as a result of the estimation made with the SVR algorithm, is low.

3.3 Case-3- Z-score normalization

In this case, Z-score normalization was applied to the dataset as a data preprocessing method. In the Z-score normalization process, the input features are standardized using mean and variance to create the dataset. This process does not restrict the ranges of the dataset. The performance evaluation metrics of the five different ML algorithms that were trained and tested, namely RMSE, R^2, MAE, MSE values, as well as the prediction speed and training time values, are presented in Table 6.

When the obtained results were examined, the lowest RMSE value was obtained in the SVR algorithm with processed data. The RMSE value, which was 0.41672 in the raw data set, improved the performance of the algorithm when the data was processed and decreased to 0.28928. In addition, the R^2 value, which was 0.95354 in the raw data set,

Table 5.
Comparison of different ML regression algorithm performance for case-2.

Regression algorithm	Data- processing	RMSE	MSE	R-square (R^2)	MAE	Prediction speed	Training time
LR	Min-Max	0.41708	0.17396	0.93191	0.2815	5313.3	1.3325
LSR	Min-Max	0.44072	0.19424	0.92397	0.3098	4900.9	5.9824
SVR	Min-Max	0.225	0.050626	0.98018	0.15262	7973.4	19.167
GPR	Min-Max	0.31747	0.10079	0.96055	0.17018	13206	3.3379
ANN	Min-Max	0.23816	0.056719	0.9778	0.15557	5482.1	6.556

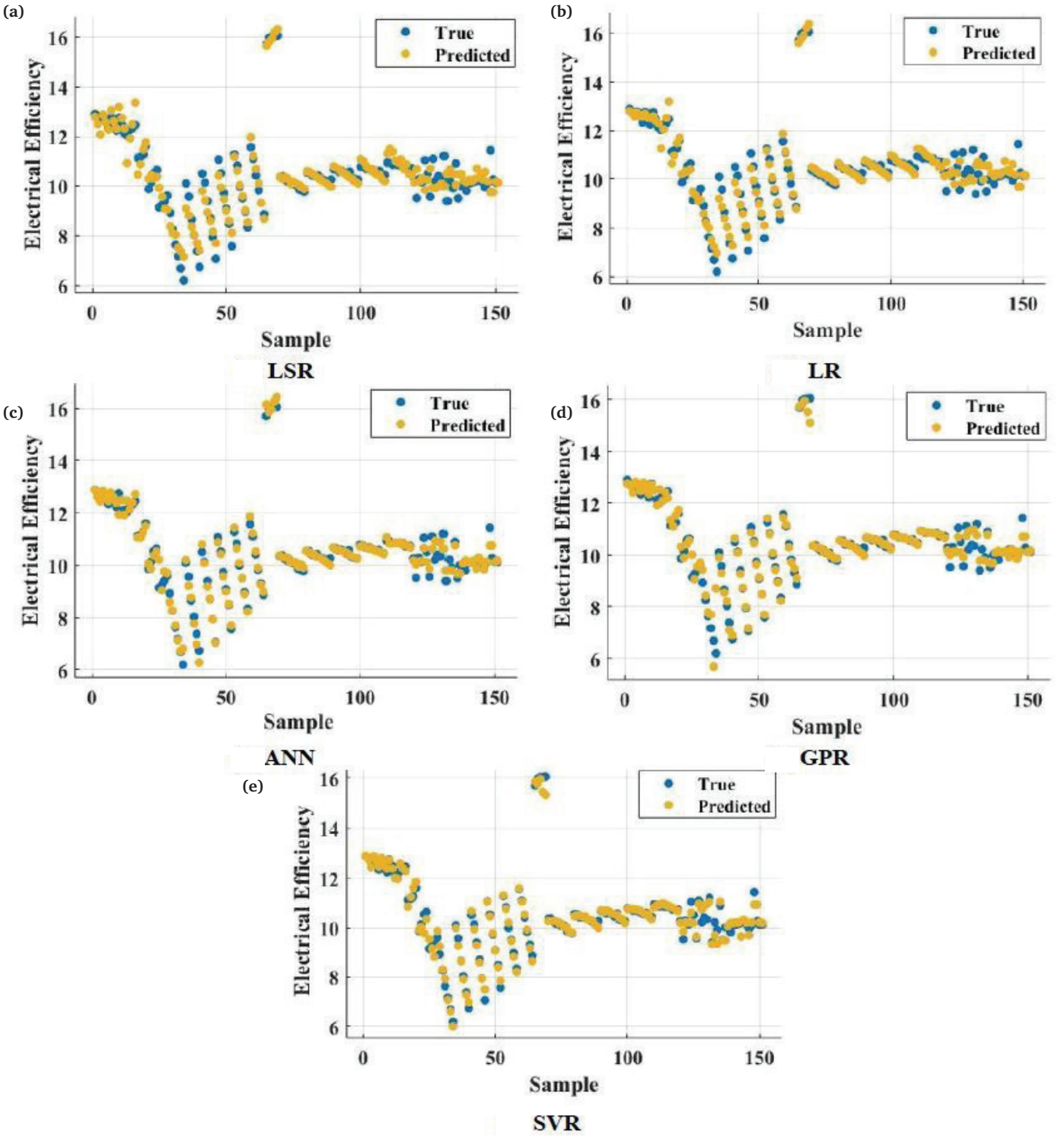


Fig. 7. Scatter plot of ML algorithms in case-2. (a) LSR, (b) LR, (c) ANN, (d) GPR, (e) SVR.

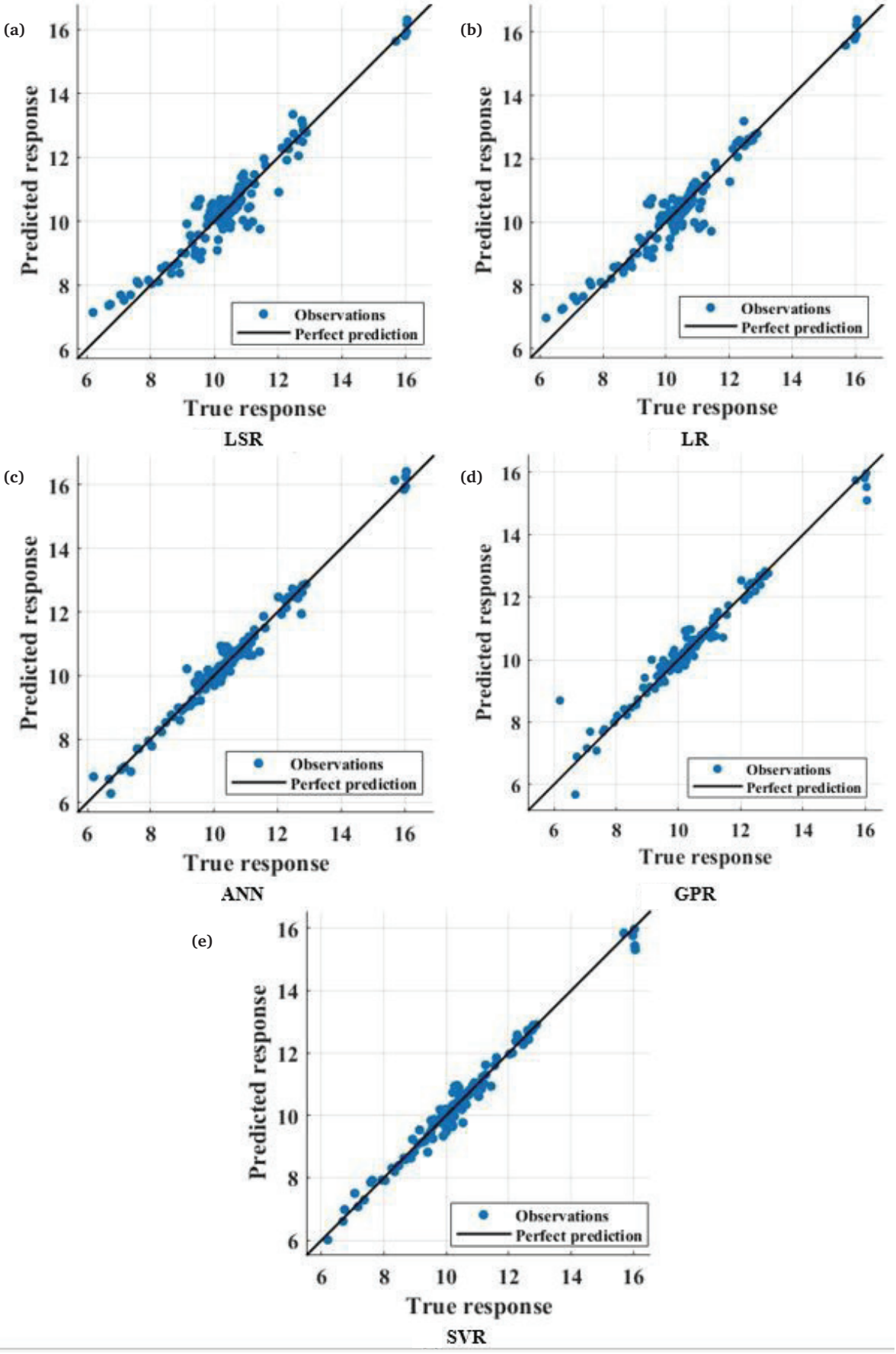


Fig. 8. Predicted and true response Scatter plot of ML algorithms in case-2. (a) LSR, (b) LR, (c) ANN, (d) GPR, (e) SVR.

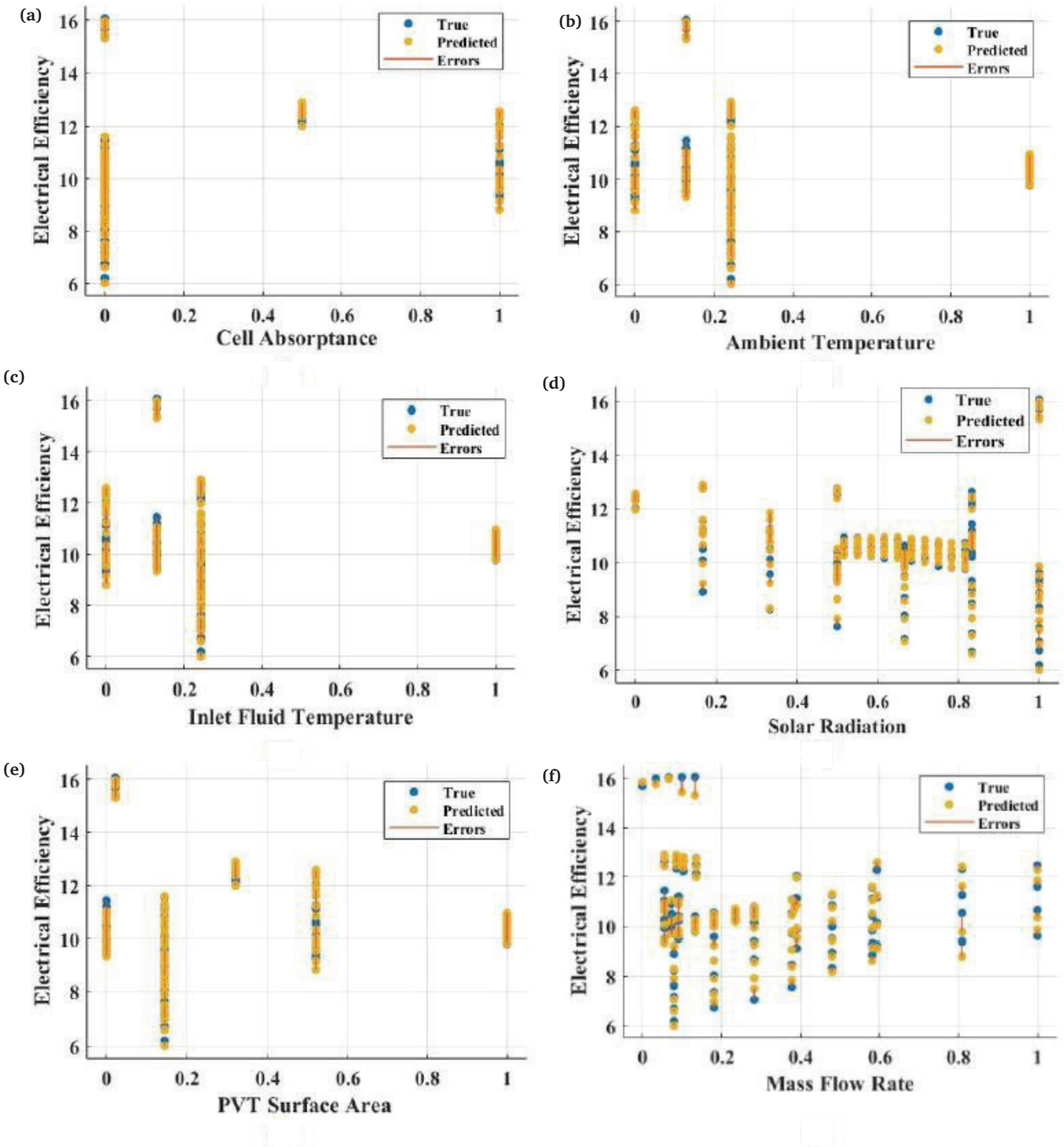


Fig. 9. True- predicted and errors values scatter graphics of different input variables of SVR in case-2. (a) Cell absorptance, (b) Ambient temperature, (c) Inlet fluid temperature, (d) Solar radiation, (e) PVT Surface area, (f) Mass flow rate.

Table 6.
Comparison of different ML regression algorithm performance for case-3.

Regression algorithm	Data- processing	RMSE	MSE	R-square (R^2)	MAE	Prediction speed	Training time
LR	Z-Score	0.41775	0.17451	0.93343	0.2779	4124.1	1.5025
LSR	Z-Score	0.43942	0.19309	0.92634	0.30396	6022.6	6.1745
SVR	Z-Score	0.28928	0.083684	0.96808	0.17489	5312.1	25.532
GPR	Z-Score	0.3206	0.10279	0.96079	0.1697	2057.2	10.007
ANN	Z-Score	0.31948	0.10207	0.96106	0.18851	6780.4	7.9653

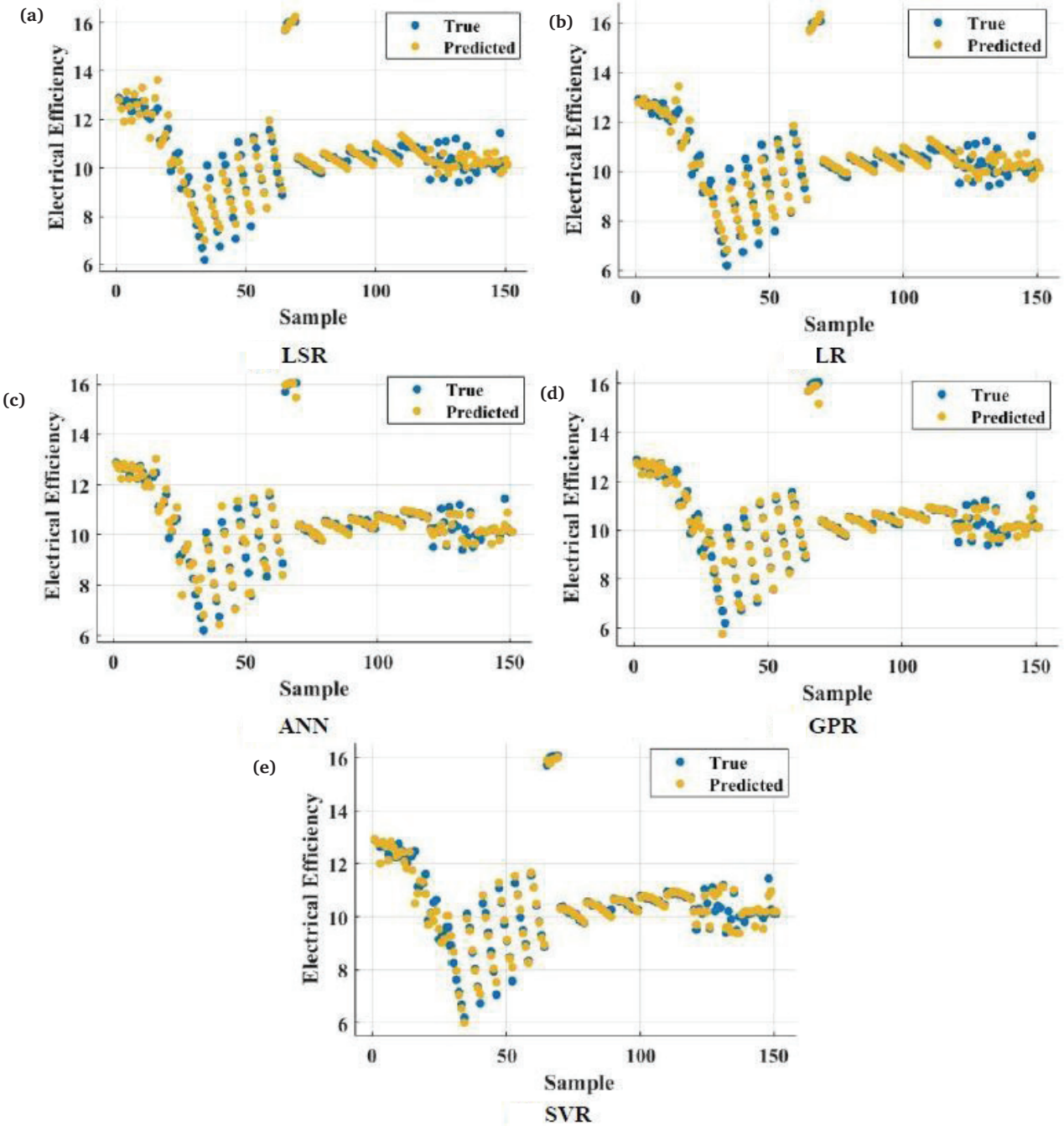


Fig. 10. Scatter plot of ML algorithms in case-3. (a) LSR, (b) LR, (c) ANN, (d) GPR, (e) SVR.

increased to 0.96808. When RMSE values and R^2 values are compared, the best performing algorithms after the SVR algorithm are ANN, GPR, LR and LSR, respectively. Although the training speed of the SVR algorithm takes longer than other algorithms, its prediction speed is 3rd with 5312.1 obs/sec.

Fig. 10 shows the graphs containing the actual and predicted data of the regression process performed with the data set that has undergone the z-Score Normalization process of the regression algorithms. In the scatter plot of the SVR algorithm, it is observed that the correct data and the predicted data overlap the most. When the graphs are examined, it is seen that the lowest performance is obtained in the LSR algorithm.

Fig. 11 presents predicted-true response graphs of regression algorithms for case-3. It is seen that the values of the predictions made in LR, LSR, ANN and GPR algorithms are more scattered around the perfect prediction line. It is seen that the predicted values in SVR algorithm are close to the perfect prediction line. This shows that SVR algorithm is the algorithm with the highest performance for the prediction process made with the data passed through z-score normalization process. The predicted electrical efficiency values for each input parameter in the SVR algorithm with the highest prediction performance are presented in Fig. 12.

3.4 Case-4 – logarithmic transform+ standardization

In this case, the raw data was first subjected to a logarithmic transformation, and then the resulting data was standardized. The logarithmic transformation aims to reduce the differences between data scales by compressing high-magnitude values and expanding low-magnitude values. The data processed in this way was used as input in five different ML algorithms to predict electrical efficiency. Evaluation

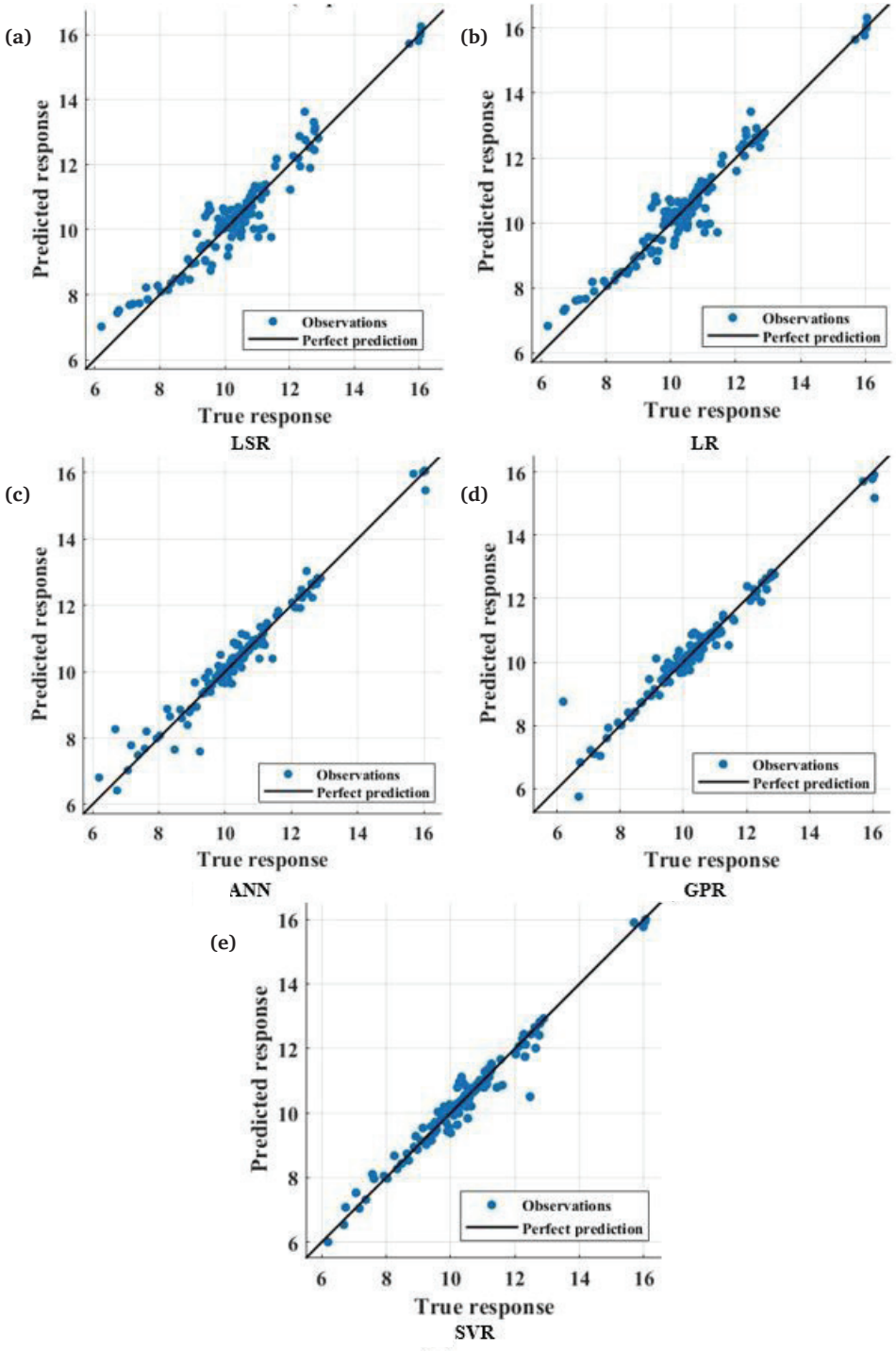


Fig. 11. Predicted and true response Scatter plot of ML algorithms in case-3.(a) LSR, (b) LR, (c) ANN, (d) GPR, (e) SVR.

metrics regarding the performance, training time, and prediction speed of the algorithms are presented in Table 7.

When the obtained results are compared, the lowest RMSE value was obtained in the LR algorithm with a value of 0.3509. The highest RMSE value was obtained in the SVR algorithm as 0.82426. When compared with the study conducted with the raw data set, it was observed that the performance of the LR and LSR algorithms increased, while the performance of the other algorithms decreased in electrical efficiency estimation. The R^2 value in the LR algorithm is the highest with 0.95226, which shows that the LR algorithm is the most suitable algorithm in electrical efficiency estimation among the five algorithms.

Fig. 13 shows the scatter plots containing the real and estimated data in the electrical efficiency prediction with the data set that has undergone logarithmic transformation + standardization process of the regression algorithms. In the scatter plots of the LR and LSR algorithms, it is seen that the actual data and the estimated efficiency values overlap more than the other algorithms. The lowest performance belongs to the SVR algorithm.

Fig. 14 presents predicted-true response graphs of regression algorithms for case-4. It is seen that the values of the predictions made in ANN, GPR and SVM algorithms are more scattered around the perfect prediction line. It is seen that the prediction data produced by LR and LSR algorithms are located near the perfect prediction line. This shows that LR and LSR algorithms are the highest performing algorithms for the prediction process made with the data passed through the log transform + standardization process. The electrical efficiency values predicted for each input parameter in the LR algorithm with the highest prediction performance are presented in Fig. 15.

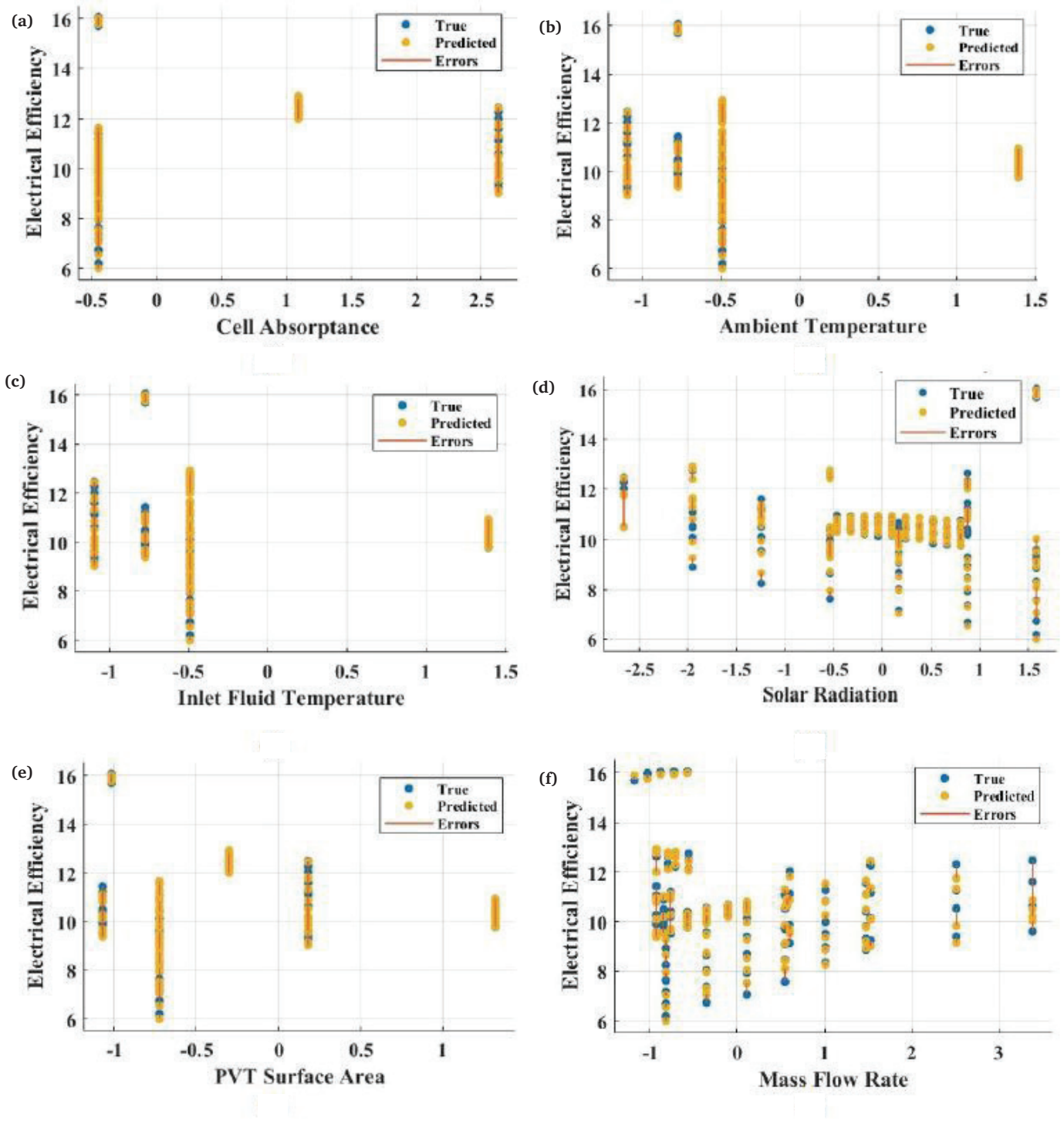


Fig. 12. True- predicted and errors values scatter graphics of Different input variables of SVR in case-3. (a) Cell absorptance, (b) Ambient temperature, (c) Inlet fluid temperature, (d) Solar radiation, (e) PVT Surface area, (f) Mass flow rate.

Performance metrics of five different ML methods (LR, LSR, ANN, SVR and GPR) were obtained in four cases where raw data and preprocessed data were used. The comparison of RMSE and R^2 values of the algorithms is presented in Table 8.

In the case of case-1, the lowest RMSE value in the electrical efficiency estimation made with the raw data set was obtained with the GPR algorithm, while in case-2, when we applied the min-max Normalization process to the data set, the lowest RMSE value was obtained with the SVR algorithm. It was observed that the performance of all regression algorithms increased with the data we applied the min-max Normalization process. This situation is also clearly seen with the increase in the R^2 value, which shows the fit of the algorithm to the problem, for all algorithms. While the R^2 value of the SVR algorithm was 0.9333 in the case where the raw data set was used, this value increased to 0.98018 with the processed data and the algorithm produced more accurate estimation values.

In case-3, z-score normalization process was applied to the data set, and the highest performance in the electrical efficiency estimation made with the obtained data belongs to the SVR algorithm with RMSE value of 0.28928. When compared with case-1 where the raw data set was used, the R^2 value of 0.9333 increased to 0.96808, increasing the algorithm's ability to adapt to the problem. It was also observed

Table 7.
Comparison of different ML regression algorithm performance for case-4.

Regression algorithm	Data- processing	RMSE	MSE	R-square (R^2)	MAE	Prediction speed	Training time
LR	Log T+Std	0.3509	0.12313	0.95226	0.22576	7186.1	1.2687
LSR	Log T+Std	0.40301	0.16242	0.93703	0.288	4965.5	6.9465
SVR	Log T+Std	0.82426	0.67941	0.73658	0.27749	9184	8.8932
GPR	Log T+Std	0.49217	0.24223	0.90608	0.23538	5116.9	2.0499
ANN	Log T+Std	0.71539	0.51179	0.80157	0.28068	6646.3	5.5986

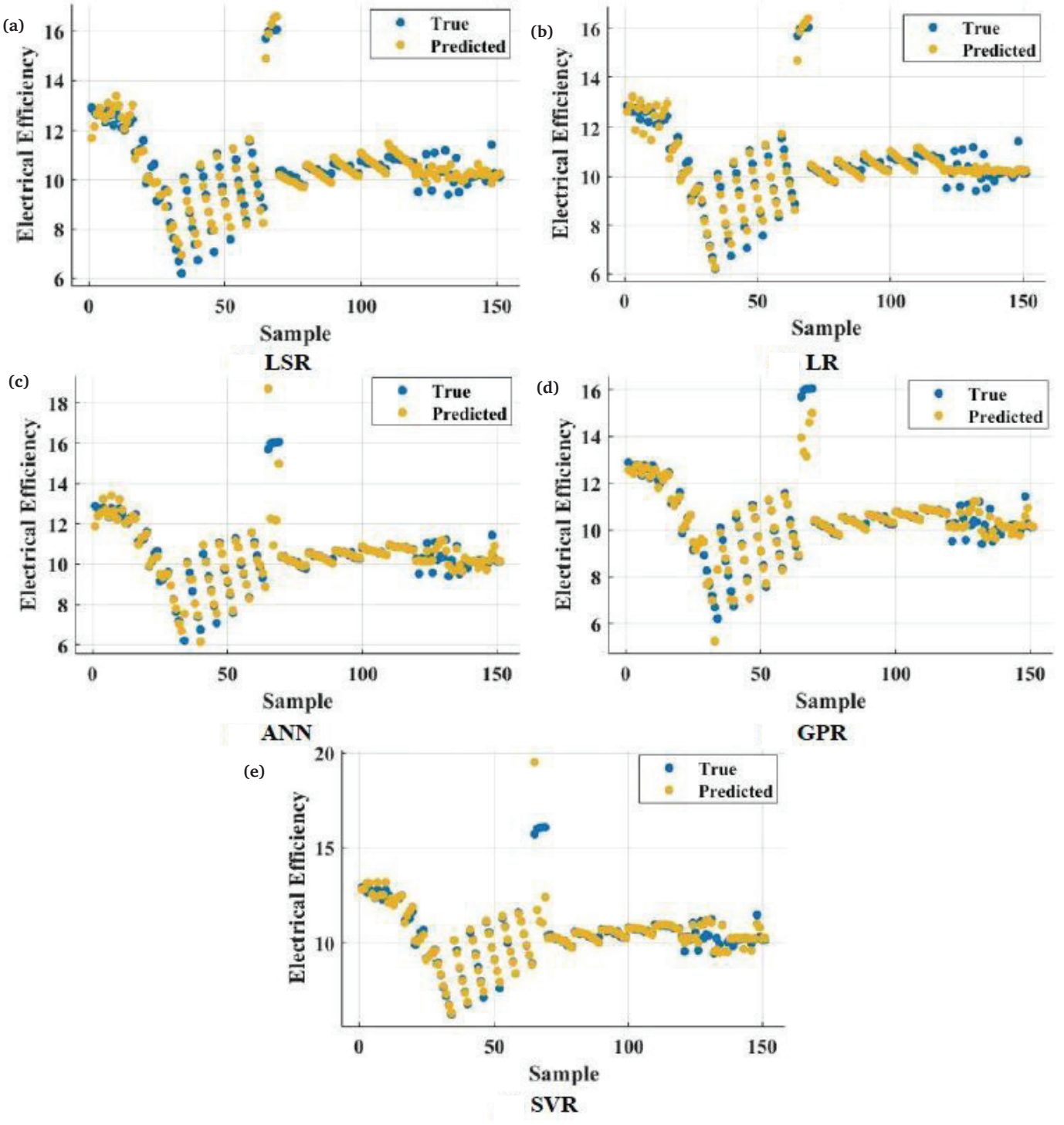


Fig. 13. Scatter plot of ML algorithms in case-4. (a) LSR, (b) LR, (c) ANN, (d) GPR, (e) SVR.

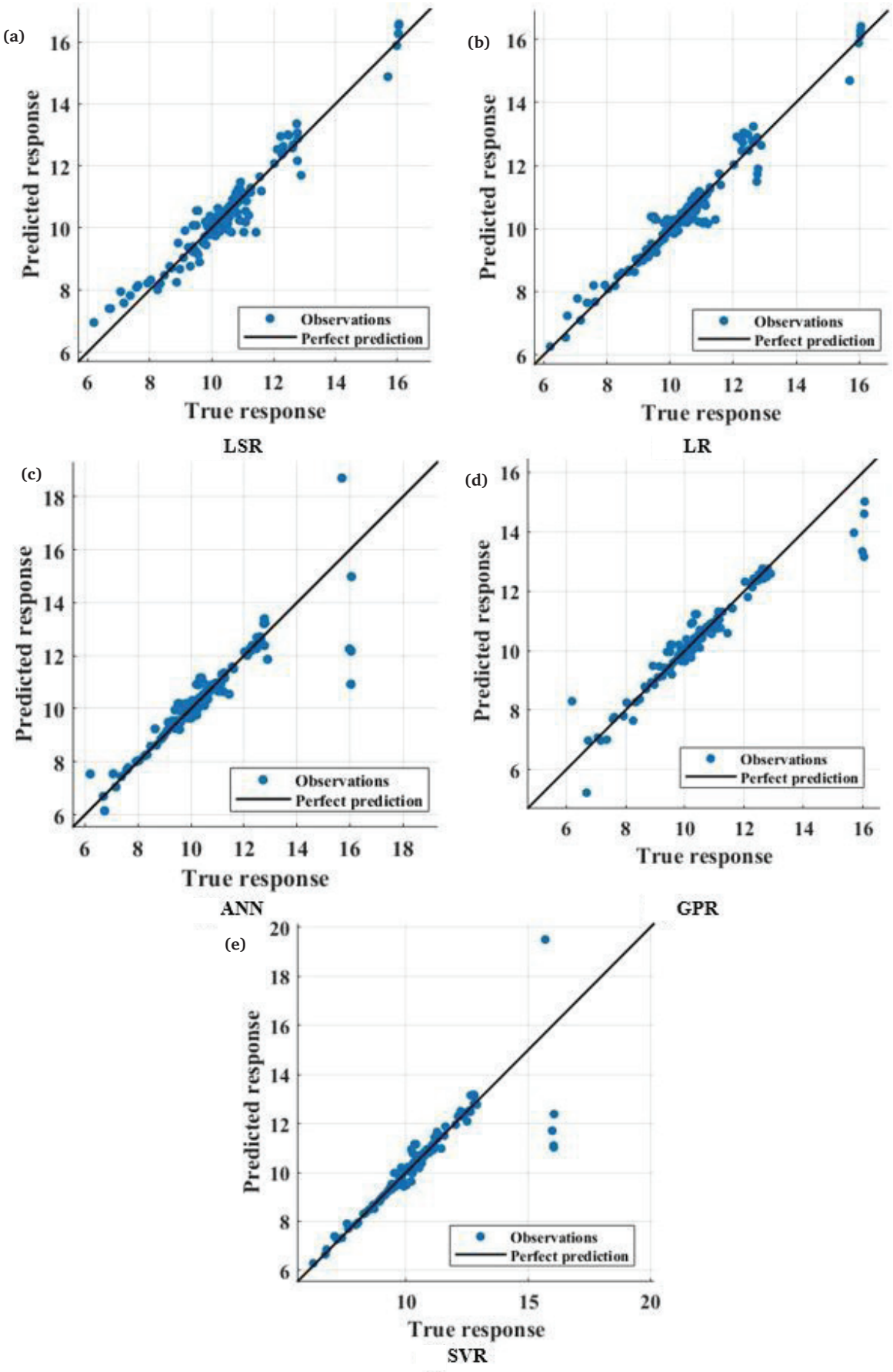


Fig. 14. Predicted and true response Scatter plot of ML algorithms in case-4. (a) LSR, (b) LR, (c) ANN, (d) GPR, (e) SVR.

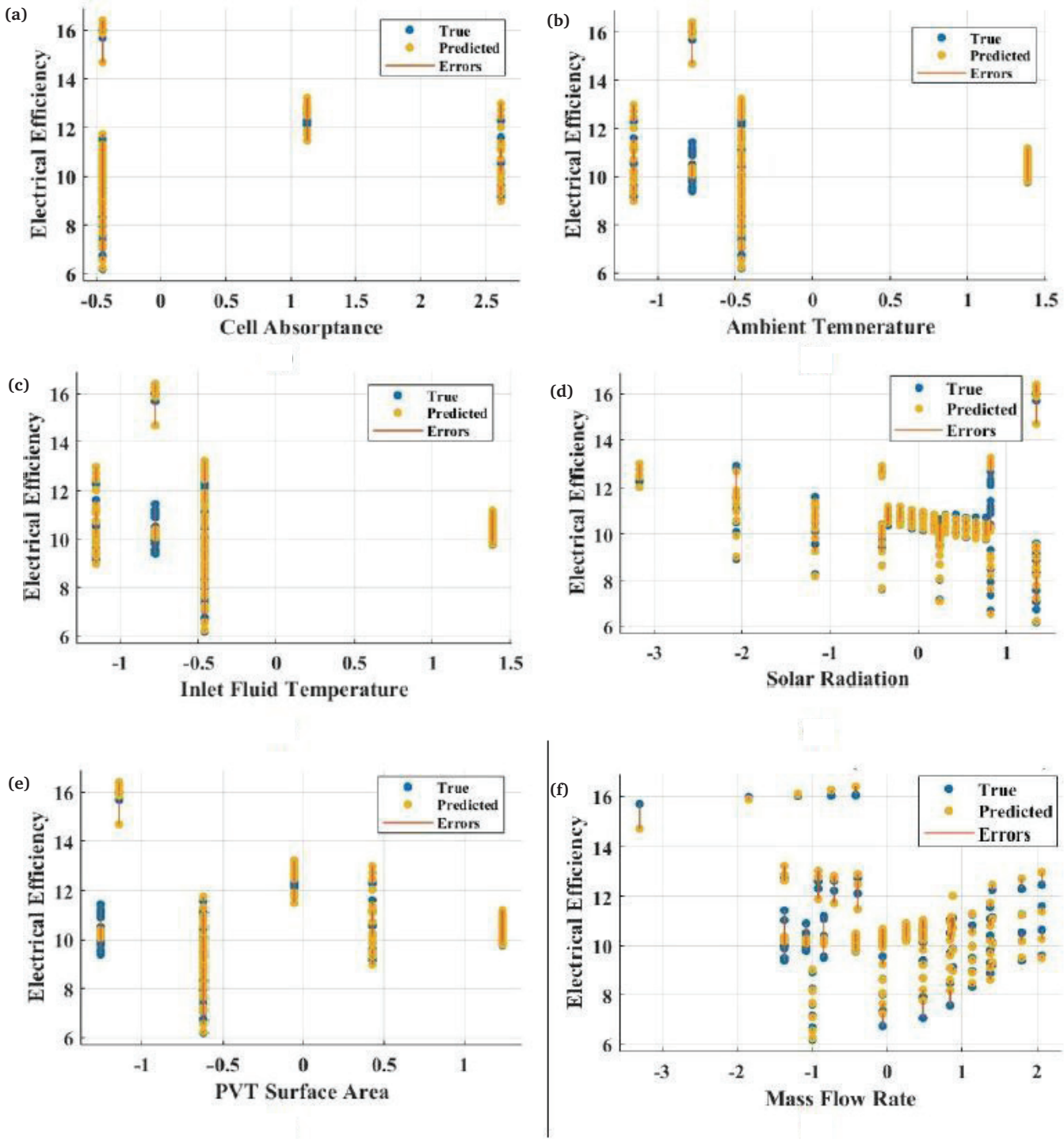


Fig. 15. True- predicted and errors values scatter graphics of different input variables of LR in case-4. (a) Cell absorptance, (b) Ambient temperature, (c) Inlet fluid temperature, (d) Solar radiation, (e) PVT Surface area, (f) Mass flow rate.

Table 8.
Comparison of RMSE and R^2 values of algorithms for all cases.

Regression algorithm	Case-1		Case-2		Case-3		Case-4	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
LR	0.42419	0,9309	0.41708	0.93191	0.41775	0.93343	0.35090	0.95226
LSR	0.45111	0.92184	0.44072	0.92397	0.43942	0.92634	0.40301	0.93703
SVR	0.41672	0.9333	0.225	0.98018	0.28928	0.96808	0.82426	0.73658
GPR	0.34778	0.95354	0.31747	0.96055	0.32060	0.96079	0.49217	0.90608
ANN	0.36561	0.94866	0.23816	0.9778	0.31948	0.96106	0.71539	0.80157

Table 9.
Comparison of results with literature.

Reference	Algorithm	RMSE	R^2
Gharaee et al. (2024)	MLP	0.7119	0.8775
	SVR	0.9884	0.7639
	RF	0.4638	0.9480
Mojumder et al. (2017)	ANN	0.347	0.8495
In this study	LR	0.35090	0.95226
	LSR	0.40301	0.93703
	GPR	0.31747	0.96055
	ANN	0.23816	0.9778
	SVR	0.225	0.98018

that the performance of all five algorithms increased. However, when case-3 is compared with case-2, it is seen that the improvement in the performance of the algorithms is lower with RMSE and R^2 values.

In case-4, logarithmic transformation + standardization process was applied to the data set. When the RMSE values of the algorithms are examined here, it is seen that the LR algorithm has the lowest value with 0.35090. When case-4 is compared with the other cases, it is seen that there is an increase in the performance of the LR and LSR algorithms, while it causes a decrease in the performance of the other three algorithms (SVR, GPR and ANN). While the processing of the data increases the adaptability of the LR and LSR algorithms to the problem (R^2 values are 0.95226, 0.93703, respectively), it decreases it in the SVR (R^2= 0.73658), GPR (R^2= 0.90608) and ANN (R^2= 0.80157) algorithms.

The results of the electrical efficiency estimation obtained in this study were compared with the literature and are presented in Table 9. Gharaee *et al.* (2024), used MLP, RF and SVR algorithms for electrical efficiency estimation in their study, and obtained testing RMSE values of 0.7119 with MLP, 0.9884 with SVR and 0.4638 with RF. In this study, the performance of the SVR method, which has a RMSE value of 0.41672, was increased by the preprocessing method applied to the data, and a RMSE value of 0.225 was obtained. When compared with the studies in the literature, it is seen that the R^2 value showing the fit of SVR to the problem among the methods proposed in this study is the best with 0.98018. Mojumder *et al.,* (2017) estimated the thermal efficiency with 0.347 RMSE value and 0.8495 R^2 value in the ANN method in their study (Mojumder *et al.,* 2017). In this study, the RMSE value of 0.36561 and R^2 value of 0.96055 obtained in the ANN method in the raw data set was reduced to 0.23816 by performing min-max normalization, and the R^2 value was increased to 0.9778, thus increasing the model's fit to the problem.

4. Conclusions

This study aims to use ML methods to overcome the uncertain situations that exist in theoretical and mathematical studies on PVT systems when the ambient conditions and parameters change. Electrical efficiency estimation was made using five algorithms (LR, LSR, SVR, GPR and ANN) from ML methods with the data collected from studies conducted for water-based PVT systems in the literature. There are 6 different input parameters in the data set collected from the literature, these are: mass flow rate, solar radiation, inlet fluid temperature, cell absorbtance, PVT surface area and ambient temperature properties. First, the performances of the algorithms were compared in the analyses performed with the raw data set, then 3 different data preprocessing methods were applied to the data set and the data was scaled and prepared for the algorithms.

The F test was applied to the data set and the effect of the input features in the data set on the prediction ability of the algorithms was obtained. The PVT surface area feature in the data set is the feature that affects the performance of the models the most. In case-1, electrical efficiency was estimated with the raw data set, and the GPR algorithm had the best performance with 0.34778 RMSE and 0.95354 R^2 value. In case-2, min-max normalization process was applied to the data, and SVR showed the best performance with 0.225 RMSE and 0.98018 R^2 value. In case-3, z-score normalization was applied to the data, and SVR showed the best performance with 0.28928 RMSE and 0.96808 R^2 value. In case-4, log transform+standardization process was applied to the data, LR showed the best performance with 0.35090 RMSE and 0.95226 R^2 value. When the 4 cases are compared, in the case where min-max normalization and z-score normalization processes are applied, an improvement in the performance of all algorithms is observed compared to case-1, where the estimation is made with the raw data set. Cross validation process is applied in order to prevent overfitting that may occur in the algorithms. Applying min-max normalization provided performance improvements across all algorithms, ensuring that each achieved the highest level of compatibility with the dataset. This preprocessing method contributed significantly to the overall model accuracy and consistency.The results show that data preprocessing methods create significant differences in the performance of the algorithms. Data preprocessing methods are used to scale the data and prepare the data for regression algorithms. By determining the appropriate preprocessing method, the prediction accuracy in the algorithms can be increased and the prediction speed can be increased.

This study compares the performance of different data preprocessing techniques and ML algorithms for estimating the electrical efficiency of PV/T systems. The results show that the Min–Max normalization + SVR model provides high accuracy and reliability among the methods examined. The proposed model can be used to reliably estimate the electrical efficiency of PV/T systems operating within the specified parameter ranges. Accurate and consistent efficiency estimates make significant contributions to feasibility analyses for early-stage design, the design and optimization of PV/T systems operating under different operating conditions, and the improvement and restructuring of existing systems. Future studies aim to increase the robustness and generalizability of the model by using larger and more diverse datasets, and to investigate different application scenarios with alternative ML algorithms.

CRediT authorship contribution statement

Merve Demirci: Conceptualized the study, developed the methodology, developed the software coding, drafted the original manuscript, investigation, review and editing of the manuscript, visualization, supervision. **Rahim Aytug Ozer:** Conceptualized the study, developed the methodology, drafted the original manuscript, formal analysis, investigation, review and editing of the manuscript, visualization, supervision.

Declaration of competing interest

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

Data availability

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors confirm that there was no use of artificial intelligence (AI)-assisted technology for assisting in the writing or editing of the manuscript and no images were manipulated using AI.

References

Abdul-Ganiyu, S., Quansah, D.A., Ramde, E.W., Seidu, R., Adaramola, M.S., 2021. Study effect of flow rate on flat-plate water-based photovoltaic-thermal (PVT) system performance by analytical technique. J Cleaner Prod 321, 128985. https://doi.org/10.1016/j.jclepro.2021.128985

Abdullah, A.L., Misha, S., Tamaldin, N., Rosli, M.A.M., Sachit, F.A., 2020. Theoretical study and indoor experimental validation of performance of the new photovoltaic thermal solar collector (PVT) based water system. Case Stud Therm Eng 18, 100595. https://doi.org/10.1016/j.csite.2020.100595

Ahmadi, M.H., Baghban, A., Sadeghzadeh, M., Zamen, M., Mosavi, A., Shamshirband, S., Kumar, R., Mohammadi-Khanaposhtani, M., 2020. Evaluation of electrical efficiency of photovoltaic thermal solar collector. Eng Appl Comput Fluid Mech 14, 545-565. https://doi.org/10.1080/19942060.2020.1734094

Ahmadinejad, M., Soleimani, A., Gerami, A., 2023. Performance enhancement of a photovoltaic thermal (PVT) system with sinusoidal fins: A quasi-transient energy-exergy analysis. Int J Green Energy 20, 978-996. https://doi.org/10.1080/15435075.2022.2131434

Al-Shamani, A.N., Sopian, K., Mat, S., Hasan, H.A., Abed, A.M., Ruslan, M.H., 2016. Experimental studies of rectangular tube absorber photovoltaic thermal collector with various types of nanofluids under the tropical climate conditions. Int J Green Energy124, 528-542. https://doi.org/10.1016/j.enconman.2016.07.052

Azad, A.K., Parvin, S., Hossain, T., 2024. Performance evaluation of nanofluid-based photovoltaic thermal (PVT) system with regression analysis. Heliyon 10, e29252. https://doi.org/10.1016/j.heliyon.2024.e29252

Bait, O., 2019. Exergy, environ–economic and economic analyses of a tubular solar water heater assisted solar still. J Cleaner Prod 212, 630-646. https://doi.org/10.1016/J.JCLEPRO.2018.12.015

Basuhaib, A., Selvaraj, J., Hasanuzzaman, M., Anjum, T., Kumar, L., 2024. Experimental investigation the effect of different operating parameters and optimized the water-based PVT system for domestic applications. Sol Energy 268, 112278. https://doi.org/10.1016/j.solener.2023.112278

Caner, M., Gedik, E., Keçebaş, A., 2011. Investigation on thermal performance calculation of two type solar air collectors using artificial neural network. Expert Syst with Applications 38, 1668-1674. https://doi.org/10.1016/j.eswa.2010.07.090

Chakraborty, A., Goswami, D., 2017. Prediction of slope stability using multiple linear regression (MLR) and artificial neural network (ANN). Arab J Geosci 10. https://doi.org/10.1007/s12517-017-3167-x

Chen, Y., Shi, R., Shu, S., Gao, W., 2013. Ensemble and enhanced PM10 concentration forecast model based on stepwise regression and wavelet analysis. Atmos Environ 74, 346-359. https://doi.org/10.1016/j.atmosenv.2013.04.002

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach Learn 20, 273-297. https://doi.org/10.1007/bf00994018

Demirci, M., Gözde, H., Taplamacioglu, M.C., 2023. Improvement of power transformer fault diagnosis by using sequential Kalman filter sensor fusion. Int J Electr Power Energy Syst 149, 109038. https://doi.org/10.1016/j.ijepes.2023.109038

Demirci, M., Saka, M., Gozde, H., Dursun, M., Taplamacioglu, M.C., 2022. Dempster shafer evidence theory application for fault diagnosis of power transformers. 2022 9th international conference on electrical and electronics engineering (ICEEE) Alanya, Turkey, pp. 40-44. https://doi.org/10.1109/iceee55327.2022.9772608

Diwania, S., Kumar, M., Kumar, R., Kumar, A., Gupta, V., Khetrapal, P., 2024. Machine learning-based thermo-electrical performance improvement of nanofluid-cooled photovoltaic–thermal system. Energy Environ 35, 1793-1817. https://doi.org/10.1177/0958305x221146947

Farahani, S.D., Zare, M.K., Alizadeh, A., 2025. Enhancing PVT/air system performance with impinging jet and porous media: A computational approach with machine learning predictions. Appl Energy 377, 124509. https://doi.org/10.1016/j.apenergy.2024.124509

Fudholi, A., Razali, N.F.M., Yazdi, M.H., Ibrahim, A., Ruslan, M.H., Othman, M.Y., Sopian, K., 2019. TiO2/water-based photovoltaic thermal (PVT) collector: Novel theoretical approach. Energy 183, 305-314. https://doi.org/10.1016/j.energy.2019.06.143

Fumo, N., Rafe Biswas, M.A., 2015. Regression analysis for prediction of residential energy consumption. Renewable and Sustainable Energy Reviews 47, 332-343. https://doi.org/10.1016/j.rser.2015.03.035

Gharaee, H., Erfanimatin, M., Bahman, A.M., 2024. Machine learning development to predict the electrical efficiency of photovoltaic-thermal (PVT) collector systems. Energy Convers Manag 315, 118808. https://doi.org/10.1016/j.enconman.2024.118808

Hamada, A., Emam, M., Refaey, H.A., Moawed, M., Abdelrahman, M.A., Elsayed, M.E.A., 2024. Identification of a different design of a photovoltaic thermal collector based on fuzzy logic control and the ARMAX model. Therm Sci Eng Prog 48, 102395. https://doi.org/10.1016/j.tsep.2024.102395

Henderi, H., 2021. Comparison of Min-Max normalization and Z-score normalization in the K-nearest neighbor (kNN) algorithm to test the accuracy of types of breast cancer. IJIIS : Int. J Inform Inform Systems 4, 13-20. https://doi.org/10.47738/ijiis.v4i1.73

Hong, S., Zhou, Z., 2012, Application of gaussian process regression for bearing degradation assessment. In 2012 6th international conference on new trends in information science, service science and data mining (ISSDM2012). IEEE, 644-648.

Hosseinzadeh, M., Sardarabadi, M., Passandideh-Fard, M., 2018. Energy and exergy analysis of nanofluid based photovoltaic thermal system integrated with phase change material. Energy 147, 636-647. https://doi.org/10.1016/j.energy.2018.01.073

Huang, J., Kaewunruen, S., 2023. Forecasting energy consumption of a public building using transformer and support vector regression. Energies 16, 966. https://doi.org/10.3390/en16020966

Jakhar, S., Paliwal, M.K., Kumar, M., 2023. Machine learning predictive models for optimal design of photovoltaic/thermal collector with nanofluids based geothermal cooling. Env Prog and Sustain Energy 42. https://doi.org/10.1002/ep.14131

Joshi, A.S., Dincer, I., Reddy, B.V., 2009. Thermodynamic assessment of photovoltaic systems. Sol Energy 83, 1139-1149. https://doi.org/10.1016/j.solener.2009.01.011

Kiwan, S.M., Khlefat, A.M., 2021. Thermal cooling of photovoltaic panels using porous material. Case Stud Therm Eng 24, 100837. https://doi.org/10.1016/j.csite.2020.100837

Ma, J., Fung, A.S., Brands, M., Abul Moyeed, O.M., Mhanna, A., Juan, N., 2021. Effects of photovoltaic/thermal (PV/T) control strategies on the performance of liquid-based PV/T assisted heat pump for space heating. Renewable Energy 172, 753-764. https://doi.org/10.1016/j.renene.2021.02.030

Ma, W., Tan, S., Hei, X., Zhao, J., Xie, G., 2016. A Prediction method based on stepwise regression analysis for train axle temperature. 2016 12th international conference on computational intelligence and security (CIS) Wuxi, China, 386-390. https://doi.org/10.1109/cis.2016.0094

Melaibari, A.A., Abu-Hamdeh, N.H., Alorfi, A.S., AL-bonsrulah, H.A.Z., Elsiddieg, A.M.A., 2024. New design for PVT system with elliptic cooling duct involving nanofluid in existence of MHD and utilizing TEG. Case Stud Therm Eng 53, 103815. https://doi.org/10.1016/j.csite.2023.103815

Mojumder, J.C., Ong, H.C., Chong, W.T., Izadyar, N., Shamshirband, S., 2017. The intelligent forecasting of the performances in PV/T collectors based on soft computing method. Renew Sustain Energy Rev72, 1366-1378. https://doi.org/10.1016/j.rser.2016.11.225

Muslim, N.H., Ghadhban, S.A., Hilal, K., 2020. Enhancement of solar photovoltaic module performance by using a water-cooling chamber for climatic conditions of Iraq. Int J Renew Energy Res (IJRER), 10, 1103-1110. https://doi.org/10.20508/ijrer.v10i3.10937.g7984

Ozer, R.A., Sahin, B., Ates, I., 2023. Effect of the acoustic manipulation on single phase immersion cooling performance of discretely heated vertical plate. Int J Heat Mass Transfer 217, 124649. https://doi.org/10.1016/j.ijheatmasstransfer.2023.124649

Özer, R.A., Şahin, B., Ateş, İ, 2025. The effect of ultrasound manipulation on boiling performance during immersion cooling. J Therm Anal Calorim 150, 15237-15250. https://doi.org/10.1007/s10973-025-14458-4

Panagiotidou, M., Aye, L., Rismanchi, B., 2020. Solar driven water heating systems for medium-rise residential buildings in urban mediterranean areas. Renewable Energy 147, 556-569. https://doi.org/10.1016/j.renene.2019.09.020

Ramos, A., Guarracino, I., Mellor, A., Alonso-Álvarez, D., Childs, P., Ekins-Daukes, N. J., Markides, C. N., 2017. Solar-thermal and hybrid photovoltaic-thermal systems for renewable heating. Grantham Institute Briefing Paper, https://doi.org/10.13140/RG.2.2.10473.29280

Salman, A.H.A., Hilal, K.H., Ghadhban, S.A., 2022. Enhancing performance of PV module using water flow through porous media. Case Stud Therm Eng 34, 102000. https://doi.org/10.1016/j.csite.2022.102000

Saloux, E., Teyssedou, A., Sorin, M., 2013. Analysis of photovoltaic (PV) and photovoltaic/thermal (PV/T) systems using the exergy method. Energy Build 67, 275-285. https://doi.org/10.1016/j.enbuild.2013.08.012

Sangeetha, M., Manigandan, S., Ashok, B., Brindhadevi, K., Pugazhendhi, A., 2021. Experimental investigation of nanofluid based photovoltaic thermal (PV/T) system for superior electrical efficiency and hydrogen production. Fuel 286, 119422. https://doi.org/10.1016/j.fuel.2020.119422

Shahsavar, A., Moayedi, H., Al-Waeli, A.H.A., Sopian, K., Chelvanathan, P., 2020. Machine learning predictive models for optimal design of building-integrated photovoltaic-thermal collectors. Int J Energy Res 44, 5675-5695. https://doi.org/10.1002/er.5323

Sharifzadeh, M., Sikinioti-Lock, A., Shah, N., 2019. Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and gaussian process regression. Renew Sustain Energy Rev108, 513-538. https://doi.org/10.1016/j.rser.2019.03.040

Sheikholeslami, M., Khalili, Z., 2024. Solar photovoltaic-thermal system with novel design of tube containing eco-friendly nanofluid. Renewable Energy 222, 119862. https://doi.org/10.1016/j.renene.2023.119862

Shi, Y., Feng, C., Yang, S., 2024. Predictive Modeling of forest fires in yunnan province: An integration of ARIMA and stepwise regression analysis. Appl Sci 14, 256. https://doi.org/10.3390/app14010256

Sopian, K., Al-Waeli, A.H.A., Kazem, H.A., 2020. Energy, exergy and efficiency of four photovoltaic thermal collectors with different energy storage material. J Energy Storage 29, 101245. https://doi.org/10.1016/j.est.2020.101245

Swese, E.E.H.A.O., Hançerlioğullari, A., 2022. Investigation of performance on photovoltaic/thermal (PV/T) system using magnetic nanofluids. Politeknik Dergisi, 25, 411-416. https://doi.org/10.2339/politeknik.1076781

Tatachar, A.V., 2021. Comparative assessment of regression models based on model evaluation metrics. Int J Innov Technol Explor Eng 8, 853-860.

Williams, C.K., Rasmussen, C.E., 2006. Gaussian processes for machine learning Cambridge, MA: MIT press. 2, 4

Yang, T., Athienitis, A.K., 2014. A study of design options for a building integrated photovoltaic/thermal (BIPV/T) system with glazed air collector and multiple inlets. Sol Energy 104, 82-92. https://doi.org/10.1016/j.solener.2014.01.049

Yao, J., Xu, H., Dai, Y., Huang, M., 2020. Performance analysis of solar assisted heat pump coupled with build-in PCM heat storage based on PV/T panel. Sol Energy 197, 279-291. https://doi.org/10.1016/j.solener.2020.01.002

Yesildal, F., Ozakin, A.N., Yakut, K., 2022. Optimization of operational parameters for a photovoltaic panel cooled by spray cooling. Eng Sci Technol Int J l 25, 100983. https://doi.org/10.1016/j.jestch.2021.04.002

Yilmaz, I.şık, Kaynar, O., 2011. Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. Expert Syst Appl 38, 5958-5966. https://doi.org/10.1016/j.eswa.2010.11.027

Yu, Y., Yang, H., Peng, J., Long, E., 2019. Performance comparisons of two flat-plate photovoltaic thermal collectors with different channel configurations. Energy 175, 300-308. https://doi.org/10.1016/j.energy.2019.03.054

Zain, A.M., Haron, H., Qasem, S.N., Sharif, S., 2012. Regression and ANN models for estimating minimum value of machining performance. Appl Math Model 36, 1477-1492. https://doi.org/10.1016/j.apm.2011.09.035

Zamen, M., Baghban, A., Pourkiaei, S.M., Ahmadi, M.H., 2019. Optimization methods using artificial intelligence algorithms to estimate thermal efficiency of PVT system. Energy Sci Eng 7, 821-834. https://doi.org/10.1002/ese3.312

Zhao, K., Popescu, S., Meng, X., Pang, Y., Agca, M., 2011. Characterizing forest canopy structure with lidar composite metrics and machine learning. Remote Sens Environ 115, 1978-1996. https://doi.org/10.1016/j.rse.2011.04.001