EXTRAPOLATIVE PREDICTION OF MAIN INJECTION FLOW RATE IN A COMMON-RAIL DIESEL INJECTOR USING SUPPORT VECTOR REGRESSION

Nguyen Xuan Khoa[1], Duong Ngoc Thiep[1],
Nguyen Tien Han[1,*]

DOI: https://doi.org/10.57001/huih5804.2026.129

ABSTRACT

This study presents a method for predicting the main injection flow rate under high-speed operating conditions (CH4) in a common-rail fuel injection system using a Support Vector Regression (SVR) model. Experimental data were collected on a dedicated injector test bench equipped with a temperature control unit to maintain the fuel temperature at 25°C, ensuring stable and repeatable test conditions. In the proposed approach, the main injection flow rates measured at low-load (CH1) and medium-load (CH2) operating modes were employed as input variables to construct the predictive model. Meanwhile, experimental data obtained directly under the high-speed condition (CH4) were used exclusively as an independent test dataset to evaluate the model's generalization and extrapolation capability under unseen operating conditions. The results demonstrate that the SVR model achieves high and stable predictive performance. For the training dataset, the model yields an R^2 of 0.9928, an RMSE of 0.0578mm^3/cycle, and an MAE of 0.0299mm^3/cycle. For the test dataset, the corresponding values are $R^2 = 0.9850$, RMSE = 0.0764mm^3/cycle, and MAE = 0.0587mm^3/cycle. These results confirm that the nonlinear relationship between main injection quantities at different operating modes can be effectively exploited even when a reduced set of input variables is used.

Keywords: *Common rail; Injection flow rate; Support Vector Regression; Artificial intelligence.*

[1]Hanoi University of Industry, Vietnam
[*]Email: hannt@haui.edu.vn
Received: 10/3/2026
Revised: 12/5/2026
Accepted: 29/5/2026

1. INTRODUCTION

In recent years, artificial intelligence (AI) and machine learning (ML) techniques have become essential tools in automotive engineering, particularly in internal combustion engine modeling, optimization, and control. ML-based approaches are widely applied to develop surrogate models, virtual sensors, and engine calibration support systems, significantly reducing experimental time and calibration costs [1, 2]. Compared with conventional empirical models, ML algorithms are capable of learning complex nonlinear relationships between multiple input and output variables across multi-operating conditions.

Among various ML algorithms, Support Vector Regression (SVR) has received considerable attention due to its strong nonlinear approximation capability through kernel functions and its structural risk minimization principle, which enhances generalization performance, especially for small- to medium-sized datasets. Liu et al. compared ANN, Random Forest, Gradient Boosting, and SVR for engine exhaust temperature prediction and reported that SVR achieved lower prediction errors and higher stability [3]. Similarly, Huaiyu Wang et al. [4] evaluated ANN, Gaussian Process Regression, and SVM for simultaneous prediction of performance and emissions in a Wankel engine, demonstrating the superiority of kernel-based models in nonlinear and extrapolation problems.

For diesel engines, Albayrak [5] developed a robust SVR model to predict combustion performance and emission characteristics under multiple injection strategies. The training and validation results showed extremely high predictive accuracy, with R^2 values close to unity and near-zero RMSE for combustion-phase parameters. This confirms that SVR is an effective tool for modeling combustion processes and has strong potential for engine control and calibration applications. Gu et al. [6] combined deep autoencoders with SVR to predict real-time fuel consumption and emissions in marine diesel engines, demonstrating that SVR can be effectively integrated into feedback control systems based on in-cylinder pressure data. Furthermore, multi-objective optimization studies combining SVM/SVR with evolutionary algorithms have been applied to common-rail diesel engines to simultaneously reduce fuel consumption and emissions, highlighting the role of SVR as an efficient surrogate model in calibration processes [7].

Recent review studies have also pointed out that ML models such as SVR, ANN, Random Forest, and XGBoost are increasingly replacing traditional regression models in engine research due to their ability to handle high-dimensional and nonlinear data [8,9]. However, for high-pressure common-rail injection systems, most ML-based studies focus on predicting emissions (NOx, soot), thermal efficiency, or fuel consumption, rather than directly predicting the main injection flow rate an essential parameter that strongly influences combustion behavior and engine performance.

Therefore, this study proposes an SVR-based approach to predict the main injection flow rate of a common-rail diesel injector under high-speed operation (CH4), using only data obtained at low- and medium-speed modes (CH1 and CH2), with the fuel temperature maintained at 25°C. This approach aims to evaluate the extrapolation capability of SVR under domain shift and to contribute a new perspective to injection quantity prediction in common-rail systems.

2. THEORETICAL BACKGROUND

2.1. Support Vector Regression (SVR)

Support Vector Regression (SVR) is a regression method derived from Support Vector Machine theory [10]. It aims to determine an optimal regression function such that the prediction error does not exceed a predefined ε-insensitive margin. Errors outside this margin are penalized through a loss function controlled by a regularization parameter C, which balances model accuracy and generalization capability.

The general SVR regression function is expressed as:

$$f(x) = w\emptyset(x) + b \quad (1)$$

where, $\emptyset(x)$ denotes a nonlinear mapping function that projects the input data into a higher-dimensional feature space, w is the weight vector, and b is the bias term. The training process is formulated as the following optimization problem:

$$min\frac{1}{2}\|w\|^2 + C\sum_{t=1}^{m}(\xi_i + \xi'_i) \quad (2)$$

where, *C is* the regularization parameter, ξ_i *and* ξ'_i are slack variables representing deviations outside the ε insensitive zone and m is the number of training samples. In this study, the SVR model was implemented using the scikit-learn library. The radial basis function (RBF) kernel was selected to capture nonlinear relationships. Prior to training, both input and output data were normalized using a standard scaler to improve numerical stability and convergence.

2.2. Performance Evaluation Metrics

To ensure the reliability and accuracy of the ML model, three commonly used evaluation metrics were employed: the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

Coefficient of Determination (R^2): R^2 represents the proportion of variance in the experimental injection flow rate explained by the predictive model. A value closer to 1 indicates better predictive accuracy.

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

where, y_i denotes the experimental value; y'_i denotes the experimental value; $\bar{y}$ is the mean value of the entire experimental dataset and n is the total number of data samples.

Mean Absolute Error (MAE): MAE measures the average absolute difference between the experimental injection quantity and the predicted value, regardless of the direction of the error. It provides an intuitive metric for evaluating the average deviation of the predictive model.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y'_i| \quad (4)$$

Root Mean Square Error (RMSE): RMSE represents the square root of the mean of the squared differences between the experimental and predicted injection quantities. A smaller RMSE indicates higher prediction accuracy of the model.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y'_i)^2} \quad (5)$$

where, y_i is the experimental value and y'_i is the predicted value.

3. EXPERIMENTAL SETUP AND DATA DESCRIPTION

3.1. Experimental System

Table 1. Technical specifications of the test equipment

Item	Specification
Model	CARDIV V830 (Ver. 2021)
Power supply (injector unit)	Three-phase AC, 380V, 50 - 60Hz (with integrated inverter)

Power supply (high-pressure pump)	Single-phase AC, 220V, 50 - 60Hz
Motor power	5HP (injector), 2HP (high-pressure pump)
Dimensions	1100 × 800 × 1600mm
Mass	300kg
Injector type	Bosch
Actuation	Solenoid

Table 2. Operating parameters

Parameter	Range
Test pressure	100 - 1800bar (maximum 2000bar)
Main injection duration	100 - 3000µs
Injection frequency	100 - 1800Hz
Test modes	CH_1 (low speed), CH_2 (medium speed), CH_4 (high speed).

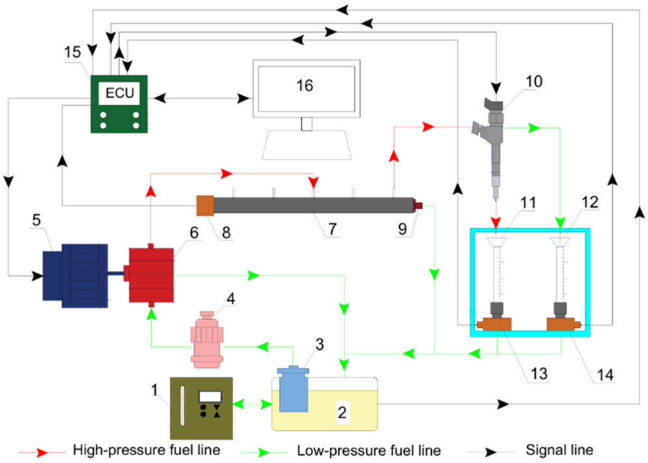

Figure 1. Experimental setup schematic

The common-rail experimental system (Figure 1) was designed as a closed-loop circulation configuration. Fuel from the tank (2) was drawn by a low-pressure pump (3) through a fuel filter (4) before being supplied to the high-pressure pump (6). Driven by an electric motor (5), the high-pressure pump compressed the fuel and delivered it to the common rail (7) at the required pressure level. The rail pressure was continuously monitored by a pressure sensor (8) and transmitted to the ECU (15), which regulated the pressure control valve (9) and the high-pressure pump to ensure stable operating conditions.

Upon receiving the control signal, the injector (10) performed the injection process. The injected fuel quantity was measured using flow sensors (11) and (12), while the return fuel flow was monitored by flow sensors (13) and (14) before returning to the fuel tank. All measurement signals were processed by the ECU and displayed on the monitoring screen (16). In addition, the system was equipped with a fuel temperature control unit (1) to maintain stable thermal conditions, thereby improving the accuracy and repeatability of the experimental results.

3.2. Dataset Description

Experimental data were collected under two operating modes: CH1 (low speed) and CH2 (medium speed), yielding a total of 232 data samples. Each sample included the following input parameters: test mode, injection frequency, main injection duration, and rail pressure. The output variable was the main injection flow rate. The dataset was randomly divided into a training set (80%, 185 samples) and a validation set (20%, 47 samples). Additionally, 128 independent samples obtained under the CH4 (high-speed) operating mode were used exclusively as a test dataset to assess the extrapolation capability of the SVR model under unseen operating conditions.

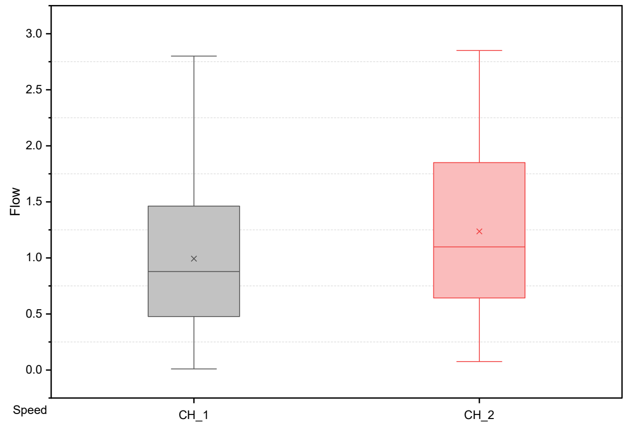

Figure 2. Distribution of input data

4. RESULTS AND DISCUSSION

The SVR model demonstrates strong extrapolation capability. Data from CH1 and CH2 were used for training and validation, while CH4 data were reserved for independent testing. The prediction results are summarized in Table 3.

Table 3. Prediction performance of the SVR model

Dataset	R^2	RMSE	MAE
Train	0.9928	0.0578	0.0299
Validation	0.9903	0.0764	0.0502
Test	0.9850	0.0875	0.0587

The results show that the SVR model achieves high and stable predictive performance across the training, validation, and test datasets, with R^2 values of 0.9928, 0.9903, and 0.9850, respectively. RMSE and MAE increase slightly from training to test (RMSE: 0.0578 - 0.0764 - 0.0875; MAE: 0.0299 - 0.0502 - 0.0587), indicating a reasonable performance degradation on unseen data without evident overfitting. The small performance gap among datasets confirms the strong generalization capability of the SVR model, consistent with the ε-insensitive loss mechanism that enhances robustness to noise in nonlinear regression. Specifically, to clearly illustrate the prediction accuracy of the SVR model, the regression plots for the training, validation, and test datasets are presented in Figure 3. The regression plots compare predicted and measured values, where the reference line Y = T represents ideal agreement. In the training set, the close clustering of points around this line indicates that the SVR model effectively captures the underlying data structure, although this mainly reflects fitting capability. The validation plot shows a similar distribution without systematic deviation, suggesting good model stability and no significant overfitting. The test plot, which is the most critical, confirms that the predicted-measured relationship remains strongly linear for independent data. The slight increase in dispersion at higher values provides insight into model sensitivity near the data boundaries and supports the model's extrapolation capability.

The residual analysis on the test dataset was performed to evaluate the error distribution, detect potential systematic bias, and assess the randomness and dispersion of prediction errors. The residuals are concentrated around zero, indicating the absence of significant systematic bias. Most errors fall within the range of –0.10 to +0.10, with a mean value close to zero, while a few extreme values (approximately +0.35 and -0.45) occur with very low frequency and can be regarded as outliers. The error magnitude is small relative to the CH_4 range (0 - 3.1), with most errors below 3 - 5% of the full scale.

The residual scatter plot shows a random distribution around the zero-error line, without observable trends or heteroscedasticity, confirming uniform error variance across the data range. Although slightly larger deviations appear at higher CH_4 values (> 2.5), their low density indicates no systematic behavior. Overall, the residual analysis confirms that the SVR model exhibits stable accuracy, low and randomly distributed errors, and negligible bias on the test dataset. In addition, the agreement between predicted and measured flow rates is further illustrated in Figure 5, which compares both values point by point along the dataset.

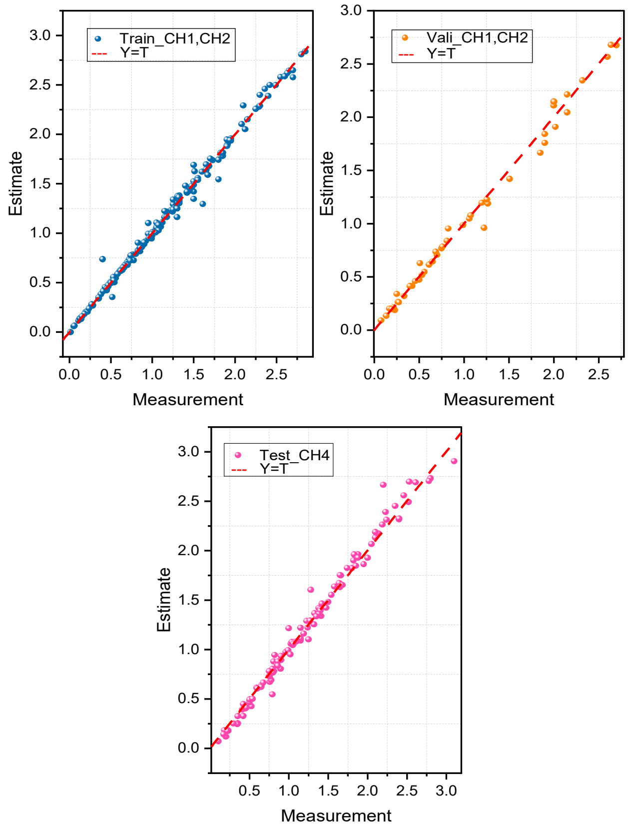

Figure 3. Regression plot between measured and predicted values

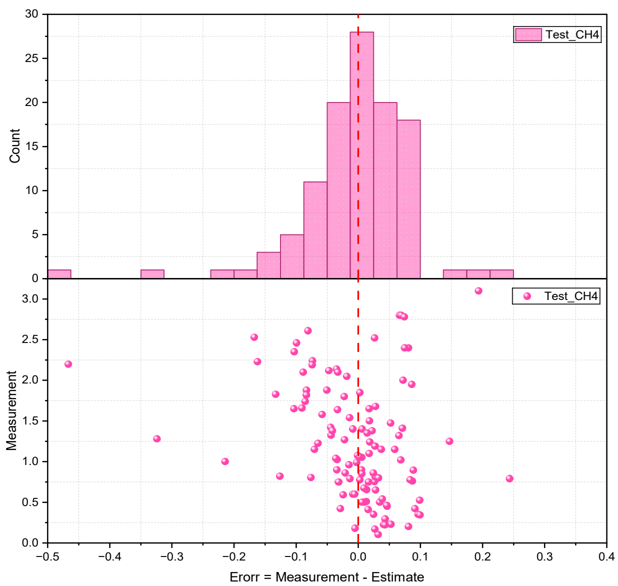

Figure 4. Distribution of CH_4 prediction errors on the test dataset

The results show that the flow rate increases progressively with the data index, from approximately 0.1 - 0.3 at the initial points to about 2.5 - 3.1 at the final points. The measured and estimated curves generally exhibit good agreement, particularly between data points ~20 and ~80, where deviations are minimal and the two curves nearly overlap. This indicates that the model accurately captures the overall variation trend of the experimental data. Local discrepancies are observed in certain regions (e.g., around points ~40 - 50 and ~90 - 100), where the estimated curve shows slightly stronger fluctuations or minor deviations from the measured values. However, these differences remain small relative to the overall scale (approximately 0 - 3.2), indicating acceptable prediction errors under highly variable conditions. Overall, the figure demonstrates that the SVR model achieves a high level of agreement with experimental data, effectively reproducing both the global increasing trend and short-term fluctuations. Despite being applied to a completely unseen dataset, the SVR model maintains high prediction accuracy, as evidenced by the close alignment between estimated and measured curves across the entire range. This confirms the model's strong sensitivity to underlying trends and local variations, as well as its robust generalization capability without overfitting, making it suitable for nonlinear and noisy time-series prediction tasks.

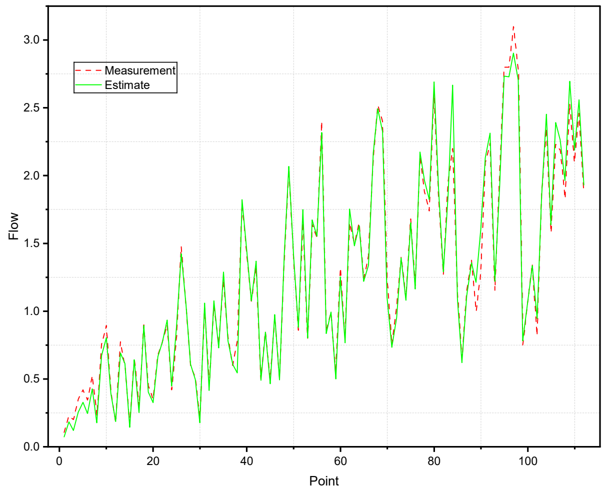

Figure 5. Comparison of measured and estimated flow rates across data points

5. CONCLUSIONS

Using an experimental dataset obtained from a common-rail fuel injection system, a Support Vector Regression (SVR) model was developed to predict the main injection flow rate under the CH4 operating mode based solely on two input variables: the main injection flow rates at CH1 and CH2. Evaluation on the test dataset demonstrates high prediction accuracy, with $R^2 = 0.9850$, RMSE = 0.0764, and MAE = 0.0587. These results indicate that the correlation among main injection quantities at different operating modes can be effectively exploited to reduce the number of input variables without compromising predictive performance. Moreover, the obtained metrics confirm the model's ability to capture nonlinear relationships between injection modes while maintaining stable performance when applied to previously unseen data.

REFERENCES

[1]. Sok R., Jeyamoorthy A., Kusaka J., "Novel virtual sensors development based on machine learning combined with convolutional neural-network image processing-translation for feedback control systems of internal combustion engines," *Applied Energy*, 365, 123224, 2024. https://doi.org/10.1016/j.apenergy.2024.123224

[2]. Aliramezani M., Koch C. R., Shahbakhti M., "Modeling, diagnostics, optimization, and control of internal combustion engines via modern machine learning techniques: A review and future directions," *Progress in Energy and Combustion Science*, 88, 100967, 2022. https://doi.org/10.1016/j.pecs.2021.100967

[3]. J. Liu, Q. Huang, C. Ulishney, C. E. Dumitrescu, "Machine learning assisted prediction of exhaust gas temperature of a heavy-duty natural gas spark ignition engine," *Applied Energy*, 300, 117413, 2021. https://doi.org/10.1016/j.apenergy.2021.117413

[4]. H. Wang, C. Ji, C. Shi, Y. Ge, H. Meng, J. Yang, K. Chang, S. Wang, "Comparison and evaluation of advanced machine learning methods for performance and emissions prediction of a gasoline Wankel rotary engine," *Energy*, 248, 123611, 2022. https://doi.org/10.1016/j.energy.2022.123611

[5]. Q. Wang, R. Yang, X. Sun, Z. Liu, Y. Zhang, J. Fu, R. Li, "The engine combustion phasing prediction based on the support vector regression method," *Processes*, 10, 4, 717, 2022. https://doi.org/10.3390/pr10040717

[6]. J. Gu, Y. Wang, J. Hu, K. Zhang, L. Shi, K. Deng, "Real-time prediction of fuel consumption and emissions based on deep autoencoding support vector regression for cylinder pressure-based feedback control of marine diesel engines," *Energy*, 300, 131570, 2024. https://doi.org/10.1016/j.energy.2024.131570

[7]. Y. Wang, G. Wang, G. Yao, Q. Shen, X. Yu, S. He, "Combining GA-SVM and NSGA-III multi-objective optimization to reduce the emission and fuel consumption of high-pressure common-rail diesel engine," *Energy*, 278, 127965, 2023. https://doi.org/10.1016/j.energy.2023.127965

[8]. Kumar S., Sharma P., Pal K., "Application of Machine Learning Approach in Internal Combustion Engine: A Comprehensive Review," in: Kumar, A., Zunaid, M., Subramanian, K.A., Lim, H. (eds) *Recent Advances in Manufacturing and Thermal Engineering* (RAMMTE 2022,) *Lecture Notes in Mechanical Engineering*, Springer, Singapore, 2023. https://doi.org/10.1007/978-981-19-8517-1_12

[9]. H. A. Hameed Al-Hamzawi, A. S. A. Al Sailawi, R. Z. Homod, H. A. Y. Kadhim, H. I. Mohammed, M. A. Al-Shareeda, "Machine learning-enhanced optimization of exhaust gas recirculation strategies for superior diesel engine performance and emissions control: A synergistic experimental and computational study," *International Journal of Hydrogen Energy*, 169, 151184, 2025. https://doi.org/10.1016/j.ijhydene.2025.151184