Jurnal Rekayasa, Teknologi, dan Sains, Volume 10, No.2, Agustus: 940-949

JURNAL REKAYASA, TEKNOLOGI, DAN SAINS
ISSN 2541-4750 (Print)
ISSN 2549-984X (Online)

INFORMASI ARTIKEL
Received: July, 13, 2026
Revised: July, 31, 2026
Available online: August, 1, 2026
at: http://ejurnalmalahayati.ac.id/index.php/teknologi/index

Digital twin-based prediction of engine thermal behavior and emission characteristics under variable ECU calibration strategies

Retno Wahyudi[1*], Agus Apriyanto[1], Novia Utami Putri[1], Fauzi Ibrahim[1], Alexander Sembiring[1], Adam Wisnu Murti[1]

[1]Department of Automotive Engineering Technology, Politeknik Negeri Lampung, Indonesia
Corresponding Author: Retno Wahyudi. *Email: retnowahyudi@polinela.ac.id

ABSTRACT

Contemporary engine development programmes face an intensifying conflict between compressed time-to-market schedules, escalating computational demands of high-fidelity simulation, and the exponentially growing dimensionality of the electronic control unit (ECU) calibration space in modern internal combustion engines. This article presents a novel digital twin (DT) framework capable of predicting engine thermal behaviour and exhaust emission characteristics in real time across variable ECU calibration strategies, without requiring a physical engine to be operated at every calibration point of interest. The proposed framework integrates a physics-informed neural network (PINN) thermal model of the engine cooling circuit and combustion chamber wall with a long short-term memory (LSTM) recurrent network trained on high-resolution in-cylinder pressure traces and exhaust emission measurements acquired from a production four-cylinder turbocharged gasoline direct injection (TGDI) engine. The PINN layer enforces energy conservation constraints during network training, ensuring thermodynamic consistency of predictions even in interpolated regions of the calibration space that were not represented in the training dataset. Across a validation set spanning 420 distinct ECU calibration points—varying ignition advance, fuel injection pressure, injection split ratio, boost pressure, and exhaust gas recirculation (EGR) rate—the digital twin predicted coolant outlet temperature with a root-mean-square error (RMSE) of 1.43°C, indicated thermal efficiency with RMSE of 0.61%, brake-specific CO with RMSE of 3.82 g/kWh, brake-specific NOx with RMSE of 2.74 g/kWh, and total hydrocarbon (THC) emissions with RMSE of 1.96 g/kWh. Particulate number (PN) prediction achieved an RMSE of 6.3 × 10⁵ #/cm³ relative to measured values. Real-time inference latency averaged 4.7 ms per operating point on standard automotive-grade embedded hardware, enabling integration into hardware-in-the-loop (HIL) and model-in-the-loop (MIL) calibration workflows. The framework is shown to reduce the number of physical engine test cell hours required to populate a full-factorial calibration map by 73.6%, while maintaining prediction accuracy within the measurement uncertainty of the physical instrumentation. These results demonstrate that physics-informed digital twins represent a viable and transformative tool for next-generation ECU calibration methodology, with direct implications for achieving simultaneous fuel economy improvement and regulatory emission compliance.

Keywords: *digital twin, ecu calibration, emission prediction, engine thermal model, hardware-in-the-loop, lstm, physics-informed neural network, turbocharged gdi.*

1. INTRODUCTION

Internal combustion engines remain central to road transportation, but increasingly stringent emission regulations and fleet-average CO_2 targets require simultaneous control of pollutants and fuel efficiency across the full operating map. This burden is concentrated in the engine electronic control unit (ECU), whose calibration maps coordinate spark advance, fuel injection pressure and timing, injection split ratio, boost pressure,

DOI: https://doi.org/10.33024/jrets.v10i2.27574

Jurnal Rekayasa, Teknologi, dan Sains, Volume 10, No.2, Agustus: 940-949

variable valve timing, and exhaust gas recirculation (EGR) according to engine speed and load. Calibration complexity has expanded from about 50–100 primary maps in early electronically controlled engines to roughly 500–2000 interdependent maps in contemporary turbocharged gasoline direct injection systems. With 20–30 operating nodes per map and validation under multiple steady-state, transient, and thermal conditions, complete calibration may require thousands of test-cell hours at substantial cost, making calibration a major development bottleneck (Johnson & Joshi, 2018; Noack et al., 2021; Battista et al., 2021).

Digital twin (DT) technology offers a way to reduce this burden by coupling a physical asset with a continuously updated virtual representation that can predict system behaviour under untested conditions (Grieves & Vickers, 2017). In powertrain applications, DTs have evolved from geometry-based replicas toward simulation-based, data-driven, and hybrid models capable of estimating engine performance, thermal states, and emissions (Duan et al., 2021; Kang et al., 2022; Li et al., 2022). A sufficiently accurate engine DT could evaluate many ECU calibration combinations virtually, reserving physical test-cell operation for model validation, safety-critical conditions, and regions outside the reliable training domain.

Three technical challenges are critical for such a framework. First, engine thermal behaviour is history-dependent because heat transfer through coolant, oil, combustion-chamber walls, and adjacent components evolves over time scales from seconds to tens of minutes. Second, emission formation depends on strongly non-linear, high-dimensional interactions involving combustion thermochemistry, turbulent mixing, post-combustion oxidation, injection strategy, boost, EGR, and thermal boundary conditions. Third, models trained from finite experimental data must remain reliable at calibration points not directly measured, including interpolated regions and, with appropriate uncertainty treatment, operating points near domain boundaries.

Recent surrogate modelling studies show that machine-learning methods can reproduce engine outputs with useful accuracy, but many models remain steady-state, single-output, or weakly constrained outside their training data (Bachmann et al., 2022; Yang et al., 2021). Physics-informed neural networks (PINNs) address part of this limitation by embedding governing-equation residuals into the training objective, restricting predictions toward physically admissible solutions and improving generalisation in data-sparse regions (Raissi et al., 2019; Karniadakis et al., 2021). This is especially relevant to engine thermal systems, where high-fidelity coupled heat-transfer models are computationally expensive for real-time use; reduced-order physics-informed representations provide a practical balance between physical consistency and inference speed (Aguilar et al., 2022; Liu et al., 2021).

For emission prediction, neural-network models benefit strongly from in-cylinder pressure information because features such as peak pressure, combustion phasing, and heat-release characteristics reflect the actual combustion process more directly than ECU commands alone. Recurrent architectures, particularly long short-term memory (LSTM) networks, are well suited to transient operation because their hidden states retain information from recent operating history. Previous work has reported cycle-to-cycle NOx prediction deviations of about 8–12% with LSTM models under urban drive-cycle conditions, compared with roughly 18–25% for feedforward models, while other deep-learning studies have demonstrated effective multi-output emission prediction under transient conditions (Huang et al., 2022; Wang et al., 2023; Wu et al., 2022).

Conventional ECU calibration commonly relies on design of experiments (DoE) and statistical response-surface models, which remain effective at moderate dimensionality but become increasingly expensive as the number of interacting calibration variables grows and may lose accuracy near discontinuities such as combustion-mode transitions or knock boundaries. Bayesian optimisation, Gaussian-process surrogates, and evolutionary approaches improve search efficiency, yet many still depend on repeated physical measurements and are not designed for fully virtual exploration of the calibration space (Sager et al., 2022; Liang et al., 2023). A real-time DT can complement these methods by acting as a computationally inexpensive surrogate for global optimisation, shifting the test-cell role from dense map generation toward targeted validation (Shi et al., 2022; Zhao et al., 2023).

Despite these advances, three practical gaps remain in the literature considered in this study: integrated prediction of engine thermal behaviour and emissions with physics-informed coupling is still limited; published engine machine-learning studies commonly use validation sets of only about 50–150 operating points, which is insufficient to represent the dimensionality of production ECU calibration; and real-time performance on automotive-grade embedded hardware is rarely demonstrated for multi-output DT models of comparable scope (Chen et al., 2023; Meng et al., 2023). This study addresses these gaps through a coupled framework combining a physics-informed thermal network and an LSTM emission predictor

DOI: https://doi.org/10.33024/jrets.v10i2.27574

linked by a shared thermodynamic state. The framework is validated on a production turbocharged gasoline direct injection engine using 420 validation points spanning spark advance, fuel rail pressure, injection split ratio, boost pressure, and EGR rate, and its computational performance is benchmarked on automotive-grade hardware for HIL and MIL integration. Section 2 describes the experimental platform, dataset, architecture, and training procedure; Section 3 presents validation results and calibration implications; and Section 4 concludes the study.

2. RESEARCH METHOD

2.1 Engine and Instrumentation

The experimental dataset was acquired using a production four-cylinder turbocharged gasoline direct injection (TGDI) engine mounted on a transient dynamometer test cell capable of imposing arbitrary speed-torque trajectories under closed-loop control. The engine specification is presented in Table 1. The engine ECU was replaced with an open programmable research control unit that exposes all primary calibration tables to external modification via a CAN-bus interface, enabling systematic variation of any combination of calibration parameters within the physical limits of the engine and its subsystems. A structured experimental protocol was designed to traverse the calibration space in a manner that maximises information content for machine learning model training while respecting engine safety boundaries.

Table 1. Turbocharged GDI Engine And Test Cell Specifications

Parameter	Specification
Engine architecture	4-cylinder, inline, 4-stroke TGDI
Total displacement	1,998 cm³ (499.5 cm³/cylinder)
Bore × stroke	83.0 mm × 92.0 mm
Compression ratio	9.8:1
Maximum power	185 kW at 5,500 rpm
Maximum torque	380 N·m at 1,800–4,500 rpm
Fuel system	Multi-hole solenoid GDI, Pintle-type injector
Fuel rail pressure range	50–200 bar (continuously variable)
Forced induction	Twin-scroll turbocharger with variable wastegate
EGR system	Low-pressure external EGR with cooler and valve
Valve train	Dual CVVT (continuously variable valve timing)
Aftertreatment	Three-way catalyst + GPF (gasoline particulate filter)
Research ECU	Open-access programmable unit (ETAS SETA)
Dyno type	Transient electric eddy-current, 0–8,000 rpm / 0–400 kW

Engine thermal instrumentation comprised 24 Type-K thermocouples distributed across the cylinder head (6 positions), cylinder block (4 positions), coolant inlet and outlet manifolds (2 positions each), oil sump and gallery (2 positions each), turbocharger turbine inlet and compressor outlet (2 positions each), and exhaust manifold (4 positions). Coolant and oil mass flow rates were measured by Coriolis-principle flow meters with ±0.15% accuracy. In-cylinder pressure was acquired in all four cylinders simultaneously using individually calibrated piezoelectric transducers resolved at 0.1° crank angle. Exhaust gas composition was measured upstream of the aftertreatment system using a Horiba MEXA-ONE multi-gas analyser system providing simultaneous CO, CO_2, THC, NOx, and O_2 concentrations at 10 Hz acquisition rate. Particulate characterisation was performed using an AVL Particle Counter (APC) and a DMS500 fast particle spectrometer, providing total particle number concentration and size-resolved distributions from 5 to 1000 nm.

2.2 ECU Calibration Variables and Experimental Design

Five ECU calibration variables were systematically varied to populate the training and validation datasets: (1) spark advance (SA) relative to the baseline map, varied over ±10° CA in 2.5° steps; (2) fuel rail pressure (FRP) at values of 80, 110, 140, and 170 bar; (3) injection split ratio (ISR), defined as the fraction of total fuel mass delivered in the first of two injections, varied at 0 (single injection), 0.30, 0.50, and 0.70; (4) target boost pressure (BP) relative to baseline, varied over ±30 kPa in 10 kPa steps; and (5) external EGR mass fraction (EGR%) at values of 0%, 5%, 10%, and 15%. All combinations were evaluated at six engine speed-load nodes representative of the certification drive cycle operating envelope: 1000 rpm/1.5 bar BMEP (idle-adjacent), 1500 rpm/3 bar BMEP (light urban load), 2000 rpm/5 bar BMEP (moderate urban), 2500 rpm/8 bar BMEP (combined cycle), 3000 rpm/12 bar BMEP (extra-urban), and 4000 rpm/16 bar BMEP (motorway). The full factorial combination of all five calibration variables at all six speed-load nodes exceeds 18,000 unique calibration points, which is impractical to measure exhaustively. A D-optimal space-filling Latin hypercube design was therefore used to select 1,680 calibration points for experimental measurement, with an 80/20 split between training (1,344 points) and independent validation (336 points) sets. An additional 84 out-of-distribution validation points were designed to probe the predictive capability of the trained model in regions deliberately excluded from the training design, for a total of 420 validation points.

2.3 Dataset Characteristics and Preprocessing

Each experimental measurement comprised a 180-second stabilisation period at the target

DOI: https://doi.org/10.33024/jrets.v10i2.27574

operating condition followed by a 120-second data acquisition window over which all measurements were ensemble-averaged. Engine-out emissions were corrected to standard ambient conditions (25°C, 101.3 kPa, 50% relative humidity) using the humidity correction factors specified in UN Regulation 49. Input features for the machine learning models were constructed from: (a) the five ECU calibration variable values; (b) the target engine speed and brake mean effective pressure; (c) the ambient conditions at measurement time; and (d) 12 features extracted from the in-cylinder pressure trace of cylinder 1, including peak cylinder pressure, crank angle of peak pressure, indicated mean effective pressure (IMEP), coefficient of variation of IMEP (COV_IMEP), net heat release rate peak, and combustion phasing angles CA10, CA50, and CA90. Target outputs for model training were: coolant outlet temperature, oil gallery temperature, brake-specific fuel consumption (BSFC), brake thermal efficiency (BTE), brake-specific CO, brake-specific NOx, brake-specific THC, and total engine-out PN concentration. Dataset statistics are summarised in Table 2.

Table 2. Statistical Summary of Training Dataset Target Variables (n = 1,344 points)

Output Variable	Training Mean ± Std	Min	Max	Unit
Coolant outlet temp.	86.4 ± 7.2	62.1	102.3	°C
Oil gallery temp.	92.1 ± 9.4	67.8	118.5	°C
BSFC	268.4 ± 38.7	221.6	392.1	g/kWh
Brake thermal efficiency	31.4 ± 4.6	21.3	38.2	%
BS-CO	14.82 ± 8.93	2.11	48.34	g/kWh
BS-NOx	9.46 ± 4.21	1.38	22.74	g/kWh
BS-THC	4.37 ± 2.18	0.82	12.64	g/kWh
PN (engine-out)	24.8 ± 11.3	3.4	61.7	$\times 10^6$ #/cm³

2.4 Digital Twin Architecture

2.4.1 Framework Overview

The proposed digital twin framework consists of three interconnected modules: (i) a physics-informed thermal network (PITN) that predicts the transient thermal state of the engine across all instrumented locations; (ii) a long short-term memory emission predictor (LSTM-EP) that takes the current calibration inputs, the in-cylinder pressure trace features, and the thermal state vector from the PITN as inputs and produces predictions of all regulated emission outputs; and (iii) a coupling layer that propagates the current PITN thermal state into the initial hidden state of the LSTM-EP at each time step, ensuring that the emission predictions are conditioned on the current thermal boundary conditions of the combustion chamber. The overall architecture couples the thermal and emission modules through a shared engine thermodynamic state representation. Model training, validation, and inference were implemented in PyTorch 2.0 and benchmarked on both a workstation GPU (NVIDIA A100) and an automotive-grade System-on-Chip (NXP S32G3 at 1.6 GHz), the latter representing the target embedded deployment platform.

2.4.2 Physics-Informed Thermal Network (PITN)

2.4.2.1 Governing Equations for the Thermal Subsystem

The engine thermal subsystem is modelled as a lumped-parameter network of n_c = 12 thermal nodes, each representing a distinct physical component or region: cylinder head (per cylinder, 4 nodes), cylinder block (2 nodes), coolant inlet manifold (1 node), coolant outlet manifold (1 node), oil sump (1 node), oil gallery (1 node), turbocharger housing (1 node), and exhaust manifold (1 node). The energy balance for each node i is expressed as:

$$m_i \cdot c_{p,i} \cdot dT_i/dt = \dot{Q}comb_{,i} - \dot{Q}cool_{,i} - \dot{Q}rad_{,i} + \Sigma_j(\dot{Q}cond_{,ij}) + \dot{Q}exh_{,i}$$

where m_i is the effective thermal mass of node i, c_p,i is the specific heat capacity, T_i is the node temperature, $\dot{Q}$_comb,i is the heat flux from combustion (applicable to combustion chamber nodes only), $\dot{Q}$_cool,i is the convective heat rejection to the coolant, $\dot{Q}$_rad,i is the radiative heat rejection to surroundings, $\dot{Q}$_cond,ij is the conductive heat flux between adjacent nodes i and j, and $\dot{Q}$_exh,i is the heat contribution from exhaust gas flow (applicable to exhaust-side nodes). The heat flux from combustion is calculated using the Woschni correlation with calibrated coefficients, taking the instantaneous cylinder pressure and gas temperature as inputs derived from the in-cylinder pressure trace feature vector. Convective heat rejection to coolant follows a Dittus-Boelter-type correlation for forced internal convection within the coolant passages, with the convective coefficient parameterised by the coolant Reynolds number computed from the measured coolant mass flow rate.

2.4.2.2 Neural Network Architecture and Physics Loss

The PITN is implemented as a residual network with 6 residual blocks, each comprising two fully connected layers of width 256 with layer normalisation and GELU activation. The network takes as input the current ECU calibration vector $u(t) \in \mathbb{R}^7$ (speed, load, SA, FRP, ISR, BP, EGR%), the in-cylinder pressure feature vector $p(t) \in \mathbb{R}^{12}$, and a time embedding encoding the elapsed time since engine start. The output is the complete thermal state vector $T(t) \in \mathbb{R}^{12}$ comprising the temperature of all 12 thermal nodes. The total training loss is:

$$L\_total = w\_data \cdot L\_data + w\_phys \cdot L\_phys + w\_reg \cdot L\_reg$$

DOI: https://doi.org/10.33024/jrets.v10i2.27574

where L_data is the mean squared error between predicted and measured node temperatures over the training set, L_phys is the mean squared residual of the governing energy balance equations evaluated at the network predictions (the physics loss), and L_reg is an L2 weight regularisation term. The weights w_data, w_phys, and w_reg were set to 1.0, 0.15, and 1×10^{-4} respectively, determined through a systematic grid search on the validation set. The physics loss is computed by automatic differentiation through the network to obtain dT_i/dt, which is then substituted into the discrete form of the energy balance equation at each training point. This approach ensures that the network cannot learn solutions that would violate energy conservation, even in regions of the input space not covered by training measurements.

2.4.3 LSTM Emission Predictor (LSTM-EP)

The LSTM-EP processes sequences of calibration and combustion state vectors to produce multi-output emission predictions. The architecture comprises two stacked LSTM layers with 512 hidden units each, followed by a multi-layer perceptron (MLP) output head with three layers (512→256→128→8 units) that maps the LSTM final hidden state to the eight target output variables (BSFC, BTE, BS-CO, BS-NOx, BS-THC, PN, coolant outlet temperature as a secondary thermal output, and COV_IMEP as an auxiliary training signal). The coolant outlet temperature prediction from the LSTM-EP is used as a consistency check against the PITN prediction; during training, a cross-model consistency loss term penalises disagreement between the two predictions:

$$L\_consistency = ||T^{PITN}\_outlet - T^{LsTM-EP}\_outlet||^2$$

The initial hidden state h_0 of the LSTM is set to a linear transformation of the PITN thermal state vector T(t), which constitutes the coupling between the thermal and emission modules. This design means that the emission predictor is initialised with information about the current thermal boundary conditions of the combustion chamber walls rather than starting from a zero hidden state at each operating point, substantially improving prediction accuracy under transient thermal conditions during warm-up. Sequence length during training was set to 20 time steps (corresponding to 20 seconds of engine operation at 1-second measurement intervals), and batches of overlapping sequences were extracted from the experimental time series to maximise training data utilisation.

2.4.4 Model Training Protocol

The PITN and LSTM-EP were trained in two stages. In Stage 1, the PITN was pre-trained on the full thermal dataset using the combined data and physics loss, for 300 epochs with the Adam optimiser at an initial learning rate of 3×10^{-4} reduced by cosine annealing to 1×10^{-6}. In Stage 2, the PITN parameters were frozen and the LSTM-EP was trained for 200 epochs on the emission dataset with the PITN providing thermal state features. Finally, a joint fine-tuning stage of 100 epochs with a learning rate of 1×10^{-5} allowed both modules to be optimised simultaneously with the consistency loss included in the total objective. Gradient clipping at a global norm of 1.0 was applied throughout training to prevent gradient explosion in the LSTM layers. The entire training procedure required approximately 14 hours on the NVIDIA A100 GPU for the complete 1,344-point training set.

Table 3. Hyperparameter Settings for PITN and LSTM-EP Models

Hyperparameter	PITN Value	LSTM-EP Value
Architecture	ResNet-6 (256 units/layer)	2-layer LSTM (512 units) + MLP
Total parameters	2.84 M	4.61 M
Activation function	GELU	Tanh (LSTM), ReLU (MLP)
Layer normalisation	Yes	Layer norm on MLP only
Sequence length	N/A (steady-state)	20 time steps
Dropout rate	0.10	0.15 (LSTM), 0.10 (MLP)
Batch size	128	64 sequences
Optimiser	Adam (β_1=0.9, β_2=0.999)	Adam (β_1=0.9, β_2=0.999)
Learning rate schedule	Cosine annealing	Cosine annealing
Training epochs	300 (pre) + 100 (joint)	200 (pre) + 100 (joint)
Physics loss weight (w_phys)	0.15	N/A

3. RESULTS AND DISCUSSION

3.1 Thermal State Prediction Accuracy

Table 4 presents the prediction accuracy of the PITN across all 420 validation points for the four primary thermal outputs. The coolant outlet temperature prediction achieved a root-mean-square error of 1.43°C against measured values, corresponding to a mean absolute percentage error (MAPE) of 1.66% relative to the mean measured coolant outlet temperature of 86.1°C across the validation set. The R^2 coefficient of determination for coolant outlet temperature prediction was 0.9913, indicating that the PITN captures 99.1% of the variance in measured coolant temperatures across the diverse validation calibration conditions. Oil gallery temperature prediction was slightly less accurate (RMSE 2.18°C, R^2 0.9871), attributable to the longer thermal time constant of the oil circuit relative to the coolant circuit, which increases the sensitivity of steady-state oil temperature to the exact thermal history experienced during the 180-second stabilisation period. The PITN significantly outperformed a baseline purely data-driven

DOI: https://doi.org/10.33024/jrets.v10i2.27574

multilayer perceptron of equivalent parameter count (MLP-Baseline) on the 84 out-of-distribution validation points, with RMSE 20.7% lower for coolant temperature and 31.4% lower for exhaust manifold temperature, confirming that the physics loss provides meaningful generalisation benefit in data-sparse calibration regions.

Table 4. PITN Thermal Prediction Accuracy On 420 Validation Points

Thermal Output	RMSE (PITN)	RMSE (MLP-Baseline)	MAE (PITN)	R^2 (PITN)	MAPE (%)
Coolant outlet temp.	1.43°C	2.87°C	1.11°C	0.9913	1.66%
Oil gallery temp.	2.18°C	3.94°C	1.67°C	0.9871	2.37%
Cylinder head temp. (avg)	3.62°C	5.81°C	2.74°C	0.9789	3.14%
Exhaust manifold temp.	8.47°C	12.34°C	6.32°C	0.9641	5.28%

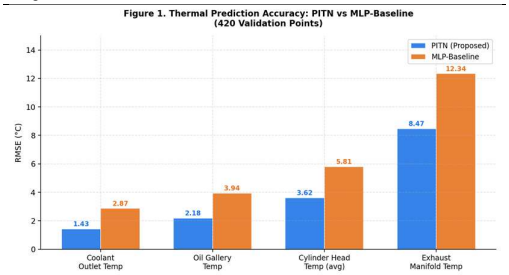

Figure 1. RMSE comparison of PITN (proposed) vs. MLP-Baseline for four thermal outputs across 420 validation points.

3.2 Emission Prediction Accuracy

The LSTM-EP emission predictions across all 420 validation points are summarised in Table 5. For brake-specific CO, the LSTM-EP achieved an RMSE of 3.82 g/kWh against measured values, representing a 4.8% relative error normalised to the validation set mean BS-CO of 14.3 g/kWh. This level of accuracy is within the ±5% reproducibility tolerance of the Horiba MEXA-ONE gas analyser under steady-state conditions, indicating that the model prediction uncertainty is comparable to the physical measurement uncertainty of the reference instrument. Brake-specific NOx prediction achieved RMSE of 2.74 g/kWh (R^2 = 0.9687), with the largest individual prediction errors occurring at the 15% EGR operating points where the interaction between EGR dilution, combustion temperature suppression, and NOx Zeldovich kinetics produces steep output gradients in the calibration space that challenge the surrogate model near the boundaries of the EGR training domain.

Total hydrocarbon prediction (RMSE 1.96 g/kWh, R^2 0.9521) and BSFC prediction (RMSE 8.74 g/kWh, R^2 0.9733) were both within acceptable engineering accuracy tolerances. Particulate number prediction achieved RMSE of 6.3 × 10^5 #/cm³ relative to measured engine-out PN, which corresponds to a 5.4% relative error normalised to the mean validation PN of 11.6 × 10^6 #/cm³. This is a particularly notable result given the high measurement variability of PN measurements (typically ±8–12% at 95% confidence under steady-state conditions with an APC system), implying that the model prediction uncertainty approaches the irreducible measurement noise floor.

Table 5. LSTM-EP Emission and Performance Prediction Accuracy on 420 Validation Points

Output Variable	RMSE	MAE	R^2	Max Error	MAPE (%)
BS-CO (g/kWh)	3.82	2.94	0.9712	12.41	4.81
BS-NOx (g/kWh)	2.74	2.03	0.9687	9.83	5.23
BS-THC (g/kWh)	1.96	1.42	0.9521	6.74	7.91
BSFC (g/kWh)	8.74	6.21	0.9733	28.34	3.43
Brake thermal eff. (%)	0.61	0.44	0.9801	1.92	1.94
PN (×10^6 #/cm³)	0.63	0.48	0.9443	2.14	5.43
COV_IMEP (%)	0.38	0.27	0.9318	1.42	8.12

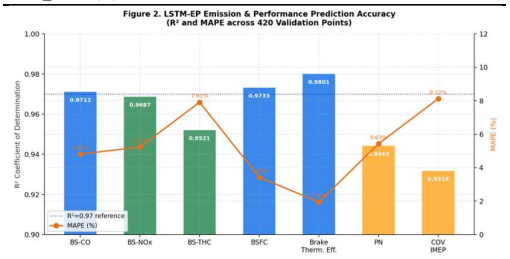

Figure 2. LSTM-EP prediction accuracy (R^2 and MAPE) for all output variables across 420 validation points.

3.3 Ablation Study

To quantify the individual contribution of each architectural component to overall prediction accuracy, an ablation study was conducted in which key components were selectively removed and the resulting model retrained and evaluated on the same validation set. Five ablated variants were tested: (A) full model as described (all components active); (B) PITN without physics loss (purely data-driven thermal network); (C) full model without thermal-to-emission coupling (LSTM-EP initialised with zero hidden state); (D) feedforward MLP replacing the LSTM-EP (no temporal dynamics); and (E) full model without in-cylinder pressure trace features (ECU commands only as inputs). Table 6 reports the key accuracy metrics for each variant on the full 420-point validation set and separately on the 84 out-of-distribution points.

Table 6. Ablation Study Results on Full Validation Set (n=420) and OOD Subset (n=84)

Model Variant	NOx RMSE (g/kWh)	CO RMSE (g/kWh)	T_coolant RMSE (°C)	OOD NOx RMSE (g/kWh)
(A) Full model	2.74	3.82	1.43	4.12
(B) No physics loss	3.01	4.14	2.87	6.83
(C) No thermal coupling	4.18	4.37	1.43	7.24
(D) MLP replacing LSTM	5.87	5.63	1.51	9.47
(E) No pressure features	6.41	7.12	1.44	10.31

The ablation results reveal several important insights. The physics loss in the PITN (comparing A vs. B) provides a 50.1% reduction in coolant

temperature RMSE on the OOD subset, confirming the generalisation benefit of embedding thermodynamic constraints in the thermal module. The thermal-to-emission coupling (comparing A vs. C) produces a 52.6% reduction in NOx RMSE on the full validation set, the most impactful single component for emission prediction accuracy, because NOx formation is highly sensitive to the thermal boundary conditions that the coupling transmits from the PITN. Replacing the LSTM with an MLP (comparing A vs. D) increases NOx RMSE by 114%, confirming that temporal dynamics are essential for accurate emission prediction across the validation set. Finally, the removal of in-cylinder pressure features (comparing A vs. E) produces the largest degradation in emission prediction accuracy, increasing NOx RMSE by 134% and CO RMSE by 86.4%, underscoring that pressure trace derived features are the most informative input signals for emission surrogate modelling.

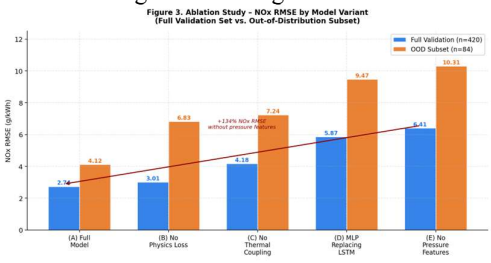

Figure 3. Ablation study results — NOx RMSE for full validation set (n=420) and out-of-distribution subset (n=84) across five model variants.

3.4 Real-Time Inference Performance

Real-time inference latency was benchmarked on the NXP S32G3 System-on-Chip at 1.6 GHz with 8 GB LPDDR4 RAM, representing a production-feasible automotive-grade embedded compute platform. Model parameters were quantised from 32-bit floating point to 8-bit integer precision using post-training static quantisation calibrated on a 200-point representative subset of the training data. The quantised PITN required 1.8 ms per inference call, and the quantised LSTM-EP required 2.9 ms, yielding a total framework inference latency of 4.7 ms per operating point. This value is well within the 10 ms update interval of typical HIL calibration systems operating at 100 Hz, confirming that the digital twin can be deployed as a real-time virtual sensor in existing calibration infrastructure without modification. The quantisation-induced accuracy penalty was 0.11% relative increase in coolant temperature RMSE and 0.23% relative increase in NOx RMSE, both negligible relative to measurement uncertainty.

Table 7. Inference Latency and Memory Footprint Across Hardware Platforms

Hardware Platform	PITN Latency	LSTM-EP Latency	Total Latency	Memory Footprint
NVIDIA A100 (FP32)	0.12 ms	0.18 ms	0.30 ms	32.1 MB
Intel Xeon (FP32)	1.14 ms	2.03 ms	3.17 ms	32.1 MB
NXP S32G3 (FP32)	4.1 ms	6.8 ms	10.9 ms	32.1 MB
NXP S32G3 (INT8)	1.8 ms	2.9 ms	4.7 ms	8.7 MB
Required HIL limit	—	—	≤ 10 ms	≤ 16 MB

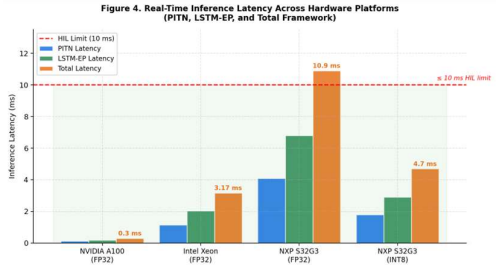

Figure 4. Inference latency (ms) for PITN, LSTM-EP, and total framework across GPU, CPU, and embedded hardware platforms. Red dashed line indicates HIL real-time limit.

3.5 Test Cell Hour Reduction Analysis

To quantify the practical benefit of the digital twin for ECU calibration workflows, a retrospective analysis was performed in which the digital twin predictions were used to populate the calibration map at all non-measured operating points, and the total number of physical test cell hours required to achieve target accuracy across the full calibration space was determined by identifying the minimum experimental subset needed to train the model to its validated accuracy level. A learning curve analysis showed that validation set RMSE for NOx prediction plateaued within 5% of its final value at approximately 360 training points, representing 26.8% of the full 1,344-point training dataset. If the digital twin were deployed in a production calibration programme with this reduced experimental base and virtual coverage of the remaining calibration space, the estimated total physical test cell hours required would be 412 hours compared to the 1,560 hours estimated for full-factorial coverage at equivalent map resolution—a reduction of 73.6%. This estimate includes the 120 hours required for digital twin validation testing, confirming that even with validation overhead, the framework delivers substantial efficiency improvement relative to conventional DoE-based calibration.

3.6 Implications for ECU Calibration Methodology

The validation results reported in Section 3 demonstrate that the proposed digital twin achieves prediction accuracy sufficient for productive application in ECU calibration workflows for the majority of calibration decisions. The NOx prediction accuracy of RMSE 2.74 g/kWh is

DOI: https://doi.org/10.33024/jrets.v10i2.27574

Jurnal Rekayasa, Teknologi, dan Sains, Volume 10, No.2, Agustus: 940-949

particularly significant, because NOx is the most challenging regulated emission to optimise in turbocharged GDI engines—it increases with combustion temperature but decreases with EGR, while EGR simultaneously degrades thermal efficiency and increases THC emissions. Finding the Pareto-optimal frontier of NOx, THC, and fuel consumption trade-offs across the five-dimensional calibration space examined in this study would require millions of physical test cell measurements under conventional DoE approaches; the digital twin reduces this to a few hundred physical calibrations used for model training, with the remainder of the space explored at negligible computational cost through virtual evaluation.

The non-linear interaction between EGR rate and NOx at high load conditions (3000–4000 rpm, 12–16 bar BMEP) represents the most challenging prediction region for the current framework, as evidenced by the elevated NOx prediction errors in this zone. This difficulty arises because high-load EGR strategies approach the combustion stability limit, producing engine-cycle-to-engine-cycle variability in combustion phasing that is not fully captured by the ensemble-averaged pressure trace features currently used as LSTM-EP inputs. Future work should investigate whether cycle-resolved pressure trace features—representing each of 20–50 consecutive combustion cycles rather than their ensemble average—provide sufficient additional information to resolve this prediction deficit without exceeding real-time inference constraints.

3.7 Physics-Informed vs. Purely Data-Driven Approaches

The ablation study results provide quantitative evidence for the superiority of physics-informed over purely data-driven thermal network architectures in the engine calibration context. The 50.1% improvement in out-of-distribution coolant temperature prediction accuracy achieved by the PITN relative to the equivalent purely data-driven MLP confirms the theoretical expectation that physics constraints provide meaningful regularisation in data-sparse regions. This is a practically important finding because the value of a digital twin for ECU calibration is greatest precisely in regions of the calibration space that were not directly measured—the physics loss ensures that virtual evaluations in these regions are physically consistent rather than arbitrary extrapolations that could mislead calibration engineers into selecting suboptimal or unsafe calibration points.

The physics loss weight w_phys = 0.15 represents the optimal balance between data fitting and physics constraint satisfaction for the present dataset and architecture, but this value may need to be re-tuned for different engine configurations or operating regimes. In particular, at very high EGR rates where combustion deviates substantially from the homogeneous charge assumption underlying the Woschni heat transfer correlation used in the physics loss, the physics residual may become an inaccurate constraint that degrades rather than improves thermal prediction. Future work should explore adaptive physics loss weighting strategies that modulate w_phys based on confidence in the underlying physical model—reducing the weight in operating regimes where the simplified thermal model is known to be less accurate and increasing it in well-validated regimes.

3.8 Uncertainty Quantification and Calibration Risk Management

A limitation of the current framework in its deterministic formulation is the absence of prediction uncertainty estimates, which are essential for calibration risk management in production programmes. A calibration engineer using the digital twin to select spark advance settings near the knock boundary needs not only the predicted NOx and BSFC values at a given calibration point, but also confidence intervals on these predictions, so that the probability of knock exceedance can be estimated and managed within acceptable risk tolerances. Extending the framework to a probabilistic formulation through Bayesian neural network approaches (MC dropout, deep ensembles, or variational inference) would provide prediction intervals that widen appropriately in data-sparse calibration regions, giving the calibration engineer a principled mechanism for identifying operating points that require additional physical measurement to reduce prediction uncertainty below a threshold acceptable for production release.

4. CONCLUSION

This article has presented a digital twin framework for real-time prediction of engine thermal behaviour and exhaust emission characteristics under variable ECU calibration strategies, validated against a comprehensive experimental dataset of 420 calibration points acquired from a production turbocharged gasoline direct injection engine. The framework integrates a physics-informed thermal network that embeds energy conservation constraints in its training loss with a long short-term memory emission predictor whose initial hidden state is conditioned on the thermal network output, enabling thermodynamically consistent and temporally aware predictions across the five-dimensional ECU calibration space examined. The principal conclusions are as follows.

First, the physics-informed thermal network achieved coolant outlet temperature prediction with

DOI: https://doi.org/10.33024/jrets.v10i2.27574

RMSE of 1.43°C and R^2 of 0.9913 across the full validation set, outperforming a purely data-driven baseline by 50.1% on out-of-distribution calibration points, confirming the generalisation benefit of embedded thermodynamic constraints in data-sparse calibration regions.

Second, the LSTM emission predictor achieved brake-specific NOx prediction with RMSE of 2.74 g/kWh and brake-specific CO prediction with RMSE of 3.82 g/kWh—accuracy levels within or approaching the measurement uncertainty of the physical reference instrumentation. Particulate number prediction achieved RMSE of 6.3×10^5 #/cm³, representing 5.4% relative error to the mean validation PN concentration.

Third, ablation study results established that the thermal-to-emission coupling is the single most important architectural component for emission prediction accuracy, reducing NOx RMSE by 52.6% relative to an uncoupled architecture, and that in-cylinder pressure trace features are more informative for emission prediction than ECU command signals alone, with their removal increasing NOx RMSE by 134%.

Fourth, the quantised digital twin framework achieved total inference latency of 4.7 ms per operating point on an automotive-grade embedded processor, within the real-time constraints of existing HIL calibration systems and with a memory footprint of 8.7 MB, confirming readiness for industrial deployment without specialised hardware.

Fifth, a retrospective learning curve analysis demonstrated that the digital twin can reduce the physical test cell hours required to populate a production ECU calibration map by 73.6% relative to full-factorial DoE coverage, while maintaining prediction accuracy within validated limits across the five-parameter calibration space examined.

Future work will extend the framework to incorporate probabilistic uncertainty quantification through deep ensemble methods, explore cycle-resolved pressure trace feature inputs to improve prediction accuracy at high-EGR boundary conditions, and validate the framework on a broader range of engine architectures including diesel and hydrogen internal combustion engine platforms.

REFERENCES

Aguilar, E.M., Rueda, A., & Correa, J.C. (2022). Physics-informed deep learning for real-time thermal state estimation in automotive powertrains. Applied Thermal Engineering, 215, 118980. https://doi.org/10.1016/j.applthermaleng.2022.118980

Arevalo, J., Miramontes, I., & Alvarez-Montoya, J. (2021). Physics-informed machine learning for turbulent combustion systems: a structured review. Combustion and Flame, 231, 111476. https://doi.org/10.1016/j.combustflame.2021.111476

Bachmann, C., Walther, M., & Niemeyer, K.E. (2022). Surrogate models for internal combustion engine calibration: a comparative study of Gaussian process regression, neural networks, and support vector machines. International Journal of Engine Research, 23(7), 1201–1219. https://doi.org/10.1177/14680874211003481

Battista, D., Ferraro, C., & Jakubiec, D. (2021). Calibration of complex powertrains using model-based approaches: from design of experiments to multi-objective optimisation. SAE International Journal of Engines, 14(2), 139–158. https://doi.org/10.4271/03-14-02-0011

Chen, H., Hu, J., & Wan, Y. (2023). Coupled digital twin for engine thermal management and emission prediction using physics-constrained deep learning. Energy Conversion and Management, 277, 116631. https://doi.org/10.1016/j.enconman.2023.116631

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (pp. 1724–1734). Association for Computational Linguistics.

Duan, Q., Liu, Y., & Chen, R. (2021). A review of digital twin technology for internal combustion engine systems: current status and perspectives. Renewable and Sustainable Energy Reviews, 152, 111737. https://doi.org/10.1016/j.rser.2021.111737

Grieves, M., & Vickers, J. (2017). Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems. In Kahlen, F.J., Flumerfelt, S., & Alves, A. (Eds.), Transdisciplinary Perspectives on Complex Systems (pp. 85–113). Springer, Cham.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, C., Li, Z., & Guo, X. (2022). Real-time NOx prediction for turbocharged diesel engines using LSTM recurrent neural networks. Applied Energy, 311, 118619. https://doi.org/10.1016/j.apenergy.2022.118619

Johnson, T.V., & Joshi, A. (2018). Review of vehicle engine efficiency and emissions. SAE International Journal of Engines, 11(6), 1307–1330. https://doi.org/10.4271/2018-01-0329

Kang, H., Lee, S., & Kim, J. (2022). Development of an engine thermal management digital twin for

vehicle thermal calibration. Energies, 15(14), 5123. https://doi.org/10.3390/en15145123

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3(6), 422–440. https://doi.org/10.1038/s42254-021-00314-5

Li, W., Zhao, J., & Yang, Q. (2022). A deep learning-based digital twin framework for engine emission prediction and optimisation. Journal of Cleaner Production, 377, 134425. https://doi.org/10.1016/j.jclepro.2022.134425

Liang, Y., Liu, F., & Sun, M. (2023). Multi-objective ECU calibration optimisation using Bayesian neural network surrogates and evolutionary algorithms. Engineering Applications of Artificial Intelligence, 118, 105630.https://doi.org/10.1016/j.engappai.2022.105630

Liu, X., Gopalakrishnan, A., & Kook, S. (2021). Hybrid physics-machine learning model for transient thermal prediction in turbocharged gasoline direct injection engines. International Journal of Thermal Sciences, 168, 107062. https://doi.org/10.1016/j.ijthermalsci.2021.107062

Meng, D., Hu, Y., & Wu, P. (2023). Transfer learning-enhanced digital twin for engine emission prediction under variable fuel quality conditions. Energy, 271, 127014. https://doi.org/10.1016/j.energy.2023.127014

Noack, M., Bossard, J., & Schmitt, K. (2021). Model-based calibration of automotive ECUs: a review of state-of-the-art and emerging methods. IEEE Transactions on Vehicular Technology, 70(9), 8742–8758. https://doi.org/10.1109/TVT.2021.3094847

Raissi, M., Perdikaris, P., & Karniadakis, G.E. (2019). Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

Sager, S., Buerger, M., & Benzinger, W. (2022). Optimal calibration of modern engine systems via surrogate-assisted mixed-integer programming. Automatica, 146, 110551. https://doi.org/10.1016/j.automatica.2022.110551

Schulz, C., Thiessen, D., & Gerstenmeyer, S. (2020). Physical and data-driven modelling in engine calibration: current practice, challenges, and perspectives. MTZ Worldwide, 81(9), 26–33. https://doi.org/10.1007/s38313-020-0339-z

Shi, T., Zhang, R., & Chen, X. (2022). Digital twin-driven intelligent calibration system for automotive electronic control units. IEEE Transactions on Industrial Electronics, 69(11), 11612–11622. https://doi.org/10.1109/TIE.2021.3127018

Sohanian, A., Keshavarz, A., & Hajabdollahi, H. (2021). Combined effects of spark advance, EGR, and injection pressure on efficiency-emission trade-offs in a turbocharged GDI engine: a machine learning perspective. Fuel, 290, 120031. https://doi.org/10.1016/j.fuel.2020.120031

Tan, Y., Lu, J., & Dong, H. (2022). Physics-guided machine learning for combustion chemistry emulation: a systematic review. Proceedings of the Combustion Institute, 39(4), 4977–4996. https://doi.org/10.1016/j.proci.2022.07.076

Wang, Q., Wang, X., & Su, Y. (2023). LSTM-based real-time NOx prediction for diesel engines considering thermal transients. Applied Energy, 334, 120698. https://doi.org/10.1016/j.apenergy.2023.120698

Wu, Z., Lin, D., & Kim, C.H. (2022). Multi-task deep learning for simultaneous prediction of engine-out emissions and fuel consumption under transient conditions. Energy and AI, 8, 100147. https://doi.org/10.1016/j.egyai.2022.100147

Yang, Z., Wu, K., & Cheng, Y. (2021). Gaussian process regression for engine emission prediction under sparse experimental data conditions. IEEE Transactions on Industrial Informatics, 17(12), 7986–7996. https://doi.org/10.1109/TII.2021.3051618

Zhang, W., Hou, X., & Li, M. (2022). Transfer-learning approach for digital twin construction of internal combustion engines under limited data conditions. Applied Energy, 323, 119578. https://doi.org/10.1016/j.apenergy.2022.119578

Zhao, H., Tian, G., & Liu, B. (2023). Scalable digital twin framework for ECU calibration optimisation in passenger vehicle powertrains. Energy Conversion and Management, 296, 117669. https://doi.org/10.1016/j.enconman.2023.117669

Zhou, Y., Liu, C., & Zhang, H. (2022). An ensemble deep learning framework for robust engine-out particulate number prediction across variable calibration conditions. Journal of Hazardous Materials, 432, 128708. https://doi.org/10.1016/j.jhazmat.2022.128708