Virtual sensor report

**Key variable summary:**

Key input: [Toque, P\_0, T\_IM, P\_IM, EGR\_Rate, ECU\_VTG\_Pos]

Potential input: [MF\_FUEL], [p\_21], [p\_31], [T\_21], [T\_31], [q\_MI], [Rail Pressure]

Output: [MF\_IA, NOx, SOC]

**Step 1: clean and prepare data**

Add processed cvs file -> only leave numerical data

Statistics summary

Must take input statistical summary

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Statistic** | **Torque** | **p0​** | **T\_IM** | **P\_IM** | **EGR\_Rate** | **ECU\_VTG\_Pos** |
| **count** | 217.000000 | 217.000000 | 217.000000 | 217.000000 | 217.000000 | 217.000000 |
| **mean** | 1129.794063 | 1.013032 | 50.199817 | 1.883465 | 23.730614 | 32.850497 |
| **std** | 681.300676 | 0.000570 | 9.196098 | 0.771776 | 8.140916 | 10.333364 |
| **min** | -0.042015 | 1.012210 | 31.000670 | 1.029631 | 3.286933 | 15.100000 |
| **25%** | 602.982100 | 1.012468 | 42.691460 | 1.209819 | 19.425220 | 24.925000 |
| **50%** | 1040.008000 | 1.013100 | 52.396680 | 1.655484 | 24.715320 | 31.004460 |
| **75%** | 1571.976000 | 1.013578 | 55.900300 | 2.394424 | 28.976458 | 40.000000 |
| **max** | 2606.458000 | 1.013860 | 69.899570 | 3.836457 | 43.376670 | 72.324840 |

Output statistical summary:

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistic** | **MF\_IA** | **NOx\_EO** | **SOC** |
| **count** | 217.000000 | 217.000000 | 217.000000 |
| **mean** | 712.738224 | 433.682002 | -2.841450 |
| **std** | 374.383776 | 318.325382 | 3.168160 |
| **min** | 146.187200 | 105.847800 | -8.179352 |
| **25%** | 424.982100 | 239.568600 | -4.972420 |
| **50%** | 632.867200 | 331.214700 | -3.588317 |
| **75%** | 979.179000 | 465.577900 | -1.027823 |
| **max** | 1632.780000 | 1579.938000 | 5.085567 |

Step 2: Explore data (data visualization)

A screenshot of a graph

AI-generated content may be incorrect.

Moderate Scatter, non-linear

Negative trend overall

Strong positive relationship (linear)

Moderate Scatter, non-linear

Very scattered spread out

positive relationship (nonlinear)

A screenshot of a graph

AI-generated content may be incorrect.

Inverse, nonlinear trend

Inverse, nonlinear trend

A screenshot of a graph

AI-generated content may be incorrect.

Inverse, nonlinear trend

Inverse, nonlinear trend

A group of green and red graphs

AI-generated content may be incorrect.A graph of a graph

AI-generated content may be incorrect.A graph with a red line

AI-generated content may be incorrect.A comparison of graphs with numbers

AI-generated content may be incorrect.

A chart of different colors

AI-generated content may be incorrect.A screenshot of a graph

AI-generated content may be incorrect.

**Model construction:**

|  |  |
| --- | --- |
| **Metric** | **Value** |
| **Total Samples** | 217 |
| **Training Set** | 173 (80%) |
| **Test Set** | 44 (20%) |
| **Key Input Features** | 6 |
| **Potential Input Features** | 7 |
| **Output Variables** | 3 |
| **Split Method** | Random stratification (seed=42) |

Simple Linear regression baseline test:

Dataset: 217 samples, 6 key inputs

**Train: 173, Test: 44 (80-20 split)**

MF\_IA: (this works best for linear regression)

------------------------------------------------------------

Train: MAE=56.1156, RMSE=68.9385, **R²=0.9655**

Test: MAE=49.9015, RMSE=63.1209, **R²=0.9729**

NOx\_EO:

------------------------------------------------------------

Train: MAE=89.0385, RMSE=119.1466, **R²=0.8423**

Test: MAE=89.9053, RMSE=130.5954, **R²=0.8780**

SOC:

------------------------------------------------------------

Train: MAE=1.2814, RMSE=1.6280, R²=0.7459

Test: MAE=1.3026, RMSE=1.6267, R²=0.6795 (slightly overfitting)

**ANN model (key input):**

Model overview:

6 key input -> 64 -> 32 -> 16 -> 1

Normalization: StandardScaler (zero mean, unit variance)

Dataset: 217 samples, 6 key inputs

Train: 173, Test: 44

MF\_IA:

------------------------------------------------------------

Train: MAE=23.1013, **RMSE**=28.4242, **R²=0.9941**

Test: MAE=22.5687, **RMSE**=28.5363, **R²=0.9945**

NOx\_EO:

------------------------------------------------------------

Train: MAE=20.5493, **RMSE**=28.5884, **R²=0.9909**

Test: MAE=29.7740, **RMSE**=39.0014, **R²=0.9891**

SOC:

------------------------------------------------------------

Train: MAE=0.1603, RMSE=0.2455, **R²=0.9942**

Test: MAE=0.2731, RMSE=0.3629, **R²=0.9841A graph with blue and red dots

AI-generated content may be incorrect.A blue squares with white text

AI-generated content may be incorrect.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F1 score** |
| **MF\_IA** | **58.3%** | **58.3%** | **58.3%** |
| **Nox\_EO** | **61.5%** | **36.4%** | **45.6%** |
| **SOC** | **65.0%** | **59.1%** | **61.9%** |

**A graph of a graph with red and blue bars

AI-generated content may be incorrect.A graph of a graph showing different colored bars

AI-generated content may be incorrect.A graph with red and green lines

AI-generated content may be incorrect.A chart with green squares

AI-generated content may be incorrect.**

overfitting

overfitting

**ANN model (both key + potential inputs):**

Model overview: 13 -> 128 -> 64 -> 32 -> 1

Dataset: 217 samples, 13 total inputs

- Key inputs: 6

- Potential inputs: 7

Train: 173, Test: 44

A graph with blue circles and red lines

AI-generated content may be incorrect.A blue squares with white text

AI-generated content may be incorrect.A graph of a graph showing different sizes of bars

AI-generated content may be incorrect.A graph of a graph showing different colored bars

AI-generated content may be incorrect.A graph with red and green lines

AI-generated content may be incorrect.A graph showing a performance metrics heatmap

AI-generated content may be incorrect.

MF\_IA:

------------------------------------------------------------

Train: MAE=12.8218, RMSE=16.3909, R²=0.9980

Test: MAE=16.1235, RMSE=20.9884, R²=0.9970

NOx\_EO:

------------------------------------------------------------

Train: MAE=15.0022, RMSE=24.7829, R²=0.9932

Test: MAE=25.1366, RMSE=35.0755, R²=0.9912

SOC:

------------------------------------------------------------

Train: MAE=0.1554, RMSE=0.2426, R²=0.9944

Test: MAE=0.2890, RMSE=0.4046, R²=0.9802

**Summary of all test results:**

MF\_IA:

Linear (6): R² ≈ 0.973, MAE ≈ 49.9 kg/h

ANN (6): R² ≈ 0.9945, MAE ≈ 22.6 kg/h

ANN (13): R² ≈ 0.9970, MAE ≈ 16.1 kg/h

NOx\_EO:

Linear (6): R² ≈ 0.878, MAE ≈ 89.9 ppm

ANN (6): R² ≈ 0.9891, MAE ≈ 29.8 ppm

ANN (13): R² ≈ 0.9912, MAE ≈ 25.1 ppm

SOC:

Linear (6): R² ≈ 0.680, MAE ≈ 1.30 deg

ANN (6): R² ≈ 0.9841, MAE ≈ 0.27 deg

ANN (13): R² ≈ 0.9802, MAE ≈ 0.29 deg (worse)

Replacement of physical sensors:

Adding extra seven physical input sensors would only yield <0.02% R^2 improvement, while significantly increases the model hidden layer compelxity, costs of buying physical sensors, and overfitting risks.

Virtual sensor (ANN- embedded in ECU)

Real-time sensor data

Sampled every engine cycle

Torque p\_0 T\_IM P\_IM EGR\_rate VTG

Physical sensor (6 key input)

Virtual sensor system

ANN model (6 -> 64 -> 32 -> 16 -> 3)

-trained on historical engine data

-runs in <1 ms per cycle

Predicted outputs

Virtual outputs (replace 3 physical sensors)

MF\_IA Nox\_EO SOC

MAE ±22.6 ±29.8 ±0.27

Financial analysis angle on replacing physical sensor:

|  |  |  |
| --- | --- | --- |
| **Aspect** | **Physical Sensors (Hardware)** | **Virtual Sensor (Model-Based)** |
| **Components Measured** | MF\_IA, NOx\_EO, SOC | MF\_IA, NOx\_EO, SOC (Derived) |
| **Additional Hardware** | Dedicated sensors/analyzers required for each (e.g., MAF meter, NOx bench analyzer, pressure transducers). | None required per vehicle (reuses existing engine sensors/ECU) OR low-cost edge box ($50–$2,000) for dedicated inference. |
| **Initial Hardware Cost** | **$1,500 – $23,000** (Varies wildly, largely driven by the cost of the high-accuracy $NOx$ analyzer). | **None** (per vehicle, if running on existing ECU) OR **$50–$2,000** for a dedicated inference unit. |
| **Development Cost** | Minimal (off-the-shelf purchase). | **One-time ≈ $3,000 – $150,000+** (Research project: $25k–$50k; Industrial-grade: $150k+). |
| **Installation Cost (Per Unit)** | **$500 – $1,500** (Labor, fittings, harnesses). | **≈ $0 – $2,000** (Software-only: firmware/ECU update; or installation of small edge unit). |
| **Installation Time** | Hours to Days. | **Minutes to Hours** (Software deployment). |
| **Calibration** | **$200 – $1,500 / vehicle** (Initial and periodic gas calibration/verification). | Minimal to None (Relies on model validation during development). |
| **Annual Maintenance** | **$100 – $2,000 / sensor / year** (High maintenance for bench analyzers; low for simple MAF). | **$500 – $5,000 /yr** (Software monitoring and periodic model retraining/updates). |
| **Lifespan** | **3 – 10 years** (Physical components degrade). | **5 – 15+ years** (Software lifespan depends on continued compatibility with ECU/hardware). |
| **5-Year Total Cost** | **≈ $3,000 – $50,000+ per unit** (The cost is multiplicative with the number of units/testbeds). | **<< Hardware alternative** when amortized over many vehicles.  one-time dev $50k amortized over 100 vehicles => ~$500/vehicle + negligible upkeep |
| **Return on Investment (ROI)** | N/A (Required capital expenditure). | **Typically 10–500×** when amortized over a large fleet, depending on the need for costly NOx hardware. |

**Neural Netwrok architecture**

**Tier 1: MLP (Primary real-time virtual sensor)**

* 6 normlazed key inputs
* Activation: ReLU
* Latency: <1 ms on modern ECU-class hardware

**Tier 2: LSTM (Temporal monitoring and drift detection)**

* Goal: detect slow drifts or degradation in inputs-outputs
* Run at lower frequency to monitor consistence between MLP predictions and temporal pattern.

**Tier 3: Ensemble of MLPs (Uncertainty quantification)**

* Quantify prediction uncertainty and increase robustness
* Trian 5-10 MLP but diff initialization

**Tier 4: Autoencoder (Sensor pattern anomaly detection)**

* Detect abnormal regimes or sensor faults
* Train an autoencoder to reconstruct normal sensor patterns. Large reconstruction error indicates an anomaly
* Raise warnings when reconstruction error exceeds a calibrated threshold

**ECU-Centric Deployment**

* **Tier 1 (MLP):** Runs every engine cycle or at fixed time intervals (e.g., 10 ms). Its outputs directly substitute MF\_IA, NOx\_EO, and SOC sensors in control or monitoring algorithms.
* **Tier 4 (Autoencoder):** Runs at a lower rate in the background to track the health of the six input sensors and detect unusual operating points.

**Optional Cloud/Backend Components**

* **Tier 2 (LSTM) and Tier 3 (Ensemble):** Can be deployed in the cloud or on an off-board server, consuming logged data from vehicles to perform deeper analysis (drift, uncertainty) and to drive periodic retraining.

**Update & Retraining Policy**

* **Monthly:** Retrain the MLP and auxiliary models using the latest aggregated field data; validate on a held-out test set.
* **Quarterly:** Roll out updated models to the fleet if they pass acceptance criteria (e.g., R² ≥ 0.985 on all outputs).
* **Annually:** Perform detailed validation vs. ground-truth lab measurements if available.