

Fault diagnostics for hydrogen-fuelled micro gas turbines using physics-based modeling and random forest classification

M Fatalla^{1,3}, A Karimi², E Saber¹, A A Taha¹ and E H Seddik¹

¹ Mechanical Engineering, Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt

² Department of Energy and Petroleum Engineering, University of Stavanger (UiS), Stavanger, Norway

³ Department of Mechanical Engineering, Al-Ryada University for Science and Technology, Sadat City, Egypt

E-mail: mustafa.m.fattalla@adj.aast.edu

Abstract. The shift toward cleaner energy systems has placed hydrogen at the forefront of fuel research, particularly for distributed power generation where Micro Gas Turbines (MGTs) offer a practical and flexible solution. A major roadblock is the lack of real-world hardware failure data for these new systems. A physics-informed framework is therefore proposed that combines a recuperated Brayton cycle thermodynamic model with a Random Forest classifier for fault identification. The hydrogen-fuelled cycle model is calibrated against six steady-state operating points from an AE-T100 test campaign, then used to generate synthetic fault datasets by simulating degradation in key components across a range of operating conditions. A Random Forest classifier is trained on this simulated data to identify fault types using only standard measurable parameters. The classifier identifies all four health states with 90.1% test accuracy (macro-F1 0.896) under a leakage-safe, run-grouped split, detecting 93.7% of faulty cases at a 5.9% false-alarm rate; when complete operating conditions are withheld from training, accuracy is 77–82% with an interpolated healthy baseline. Physics-based simulation is thus a practical way to bypass data scarcity in early-stage hydrogen systems, though field application will require testing against real degradation, sensor bias and model error.

1. Introduction

Micro gas turbines (MGTs) around 100 kW are compact, fast-starting units suited to distributed generation and combined heat and power, valued for their small footprint, low emissions, and multi-fuel capability [10]. Hydrogen firing of such machines is drawing interest as a carbon-free route to heat and power [6]: hydrogen carries about 120 MJ/kg against roughly 44 MJ/kg for gasoline, and burning it lowers the greenhouse-gas emissions that come with the fossil fuels these machines otherwise run on [11]. Hydrogen does not change the fact that every turbomachine degrades between overhauls, through mechanisms such as compressor fouling, turbine erosion, and loss of recuperator effectiveness, the heat exchanger that recovers exhaust heat and raises cycle efficiency from around 16–20% to about 30% or more in this engine class [10]. Fouling alone accounts for an estimated 70–85% of gas-turbine performance loss in service [2], and degradation of any single component shifts the operating match of

the rest of the gas path [1]. Catching and classifying that degradation from routine sensor readings, before it forces an unplanned outage, is the task of gas-path fault diagnosis.

Two things make this harder for a machine in the 100 kW class than for a large industrial turbine. The instrumentation is sparse (a few gas-path temperatures and pressures, shaft speed, fuel flow, electrical power, and nothing that measures internal component health directly), and labelled fault data barely exists, because units in the field are not deliberately run to degraded or failed states for the sake of research. The wider gas-path diagnostics literature has long recognized limitations of this kind: a recent survey lists nonlinear engine behaviour, measurement uncertainty, simultaneous faults, and a limited sensor count among the persistent obstacles for conventional diagnostic methods, and points to AI-based and hybrid techniques as the main line of response [9]. Prior MGT-specific work has dealt with the missing-data problem by building the labels from a model rather than from field failures: Talebi *et al.* combined an engine-operation database with a transformed-variable scheme to extend diagnosis across combined and part-load conditions [3], and Hashmi *et al.* paired analytical performance models of hydrogen-fuelled MGTs with neural networks for component-level fault isolation [4, 5]. The same model-then-label strategy has also been demonstrated on larger industrial turbines: injecting fouling and erosion into all six main gas-path components of a full-engine performance model produced labelled data on which k-nearest-neighbours, decision-tree, and neural-network classifiers each reached about 99% accuracy [12], evidence that a calibrated performance model can stand in for field fault data outside the MGT class as well.

Hydrogen firing adds considerations of its own. Its high flame speed and wide flammability range raise the risk of flashback and of thermal damage to burner and liner hardware, and hydrogen combustors are run lean or diluted to contain NO_x, so the combustor becomes a fault location in its own right [6, 11]. The combustion products carry much more water vapour than those of natural gas, which raises the specific heat of the hot gas and changes the turbine work and the TIT–TOT relation on which the engine controller relies; a diagnostic model and its healthy reference must therefore be evaluated for the actual fuel [4, 5]. Hydrogen's low volumetric energy density means a much larger volumetric fuel flow for the same power, which affects fuel metering, one of the few direct indicators of efficiency loss, and hydrogen-ready units are expected to run on variable hydrogen–natural-gas blends, making fuel composition an extra operating variable that can mimic or mask component degradation [6, 11]. Compared with large multi-shaft engines, the micro-turbine case is further constrained by single-shaft, variable-speed operation at constant TOT, which absorbs much of the degradation signal into the controller's own response, and by the dominant role of the recuperator, whose effectiveness sets most of the cycle efficiency [10]. This study addresses the gas-path side of the problem (compressor, turbine and recuperator faults) for pure-hydrogen operation; combustor faults and blend variability are left to future work.

This work follows the same model-then-label strategy but changes two things. The underlying model is a first-principles thermodynamic cycle model, not a black-box or purely map-based tool, and it is calibrated against a real test campaign on the specific unit being modelled rather than tuned only to plausible ranges. The classifier is trained and evaluated under a grouped train/test protocol built to keep noise realizations of the same run out of both sides of the split, a leakage mode that synthetic datasets of this kind are prone to. Section 2 covers the thermodynamic model and its calibration; Section 3, how degradation is parameterized and injected; Section 4, dataset generation; Section 5, the classifier and its results; and Section 6, the scope and limitations of the approach.

2. Thermodynamic model

The AE-T100 is a single-shaft, recuperated microturbine CHP unit with a radial centrifugal compressor (design pressure ratio approximately 4.5:1) and a firing temperature of approximately 950 °C [8]. The model reproduced here is a steady-state, station-based simulation of the recuperated Brayton cycle, implemented in MATLAB and fuelled with pure hydrogen. Figure 1 shows the cycle layout and Table 1 defines the station numbering used throughout (SAE convention); the flow path is 1 → 2 → 3 → 35 → 4 → 5 → 6, with the recuperator coupling the cold leg (3 → 35) and the hot leg (5 → 6).

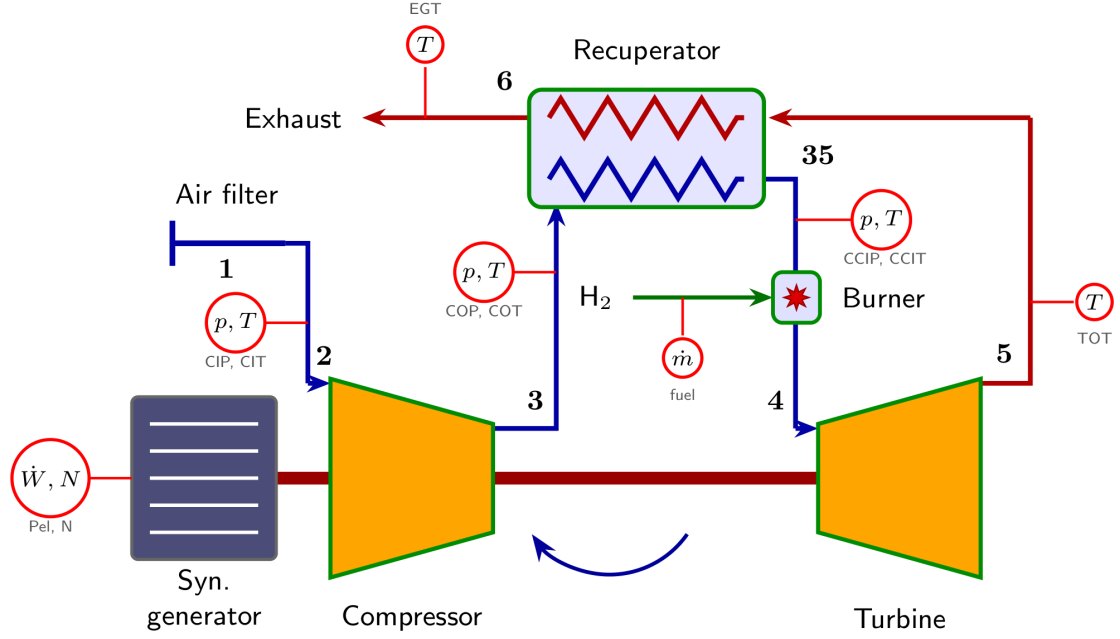


Figure 1. Station numbering and measurement layout of the recuperated, hydrogen-fuelled AE-T100. The compressed-air path (blue) runs $1 \rightarrow 2 \rightarrow 3 \rightarrow 35 \rightarrow 4$ and the hot-gas path (red) $4 \rightarrow 5 \rightarrow 6$, with the recuperator coupling the cold leg ($3 \rightarrow 35$) to the hot leg ($5 \rightarrow 6$). Circled symbols mark the eleven sensor channels available to the classifier; **station 4 (TIT) carries no instrument; the TIT used as a model input in section 2.2 is the value calculated by the unit's control system, and in prediction mode TIT is solved by the model.**

Table 1. Cycle station numbering.

Station	Location
1	Ambient
2	Compressor inlet (after inlet-duct pressure loss)
3	Compressor outlet (COT)
35	Recuperator cold-side outlet / combustor inlet (CCIT)
4	Combustor outlet / turbine inlet (TIT)
5	Turbine outlet (TOT)
6	Recuperator hot-side outlet / exhaust (EGT)

All gas properties (specific heat, enthalpy, entropy) are evaluated from NASA 7-coefficient polynomials [7] for the five species present (N_2 , O_2 , Ar , CO_2 , H_2O), rather than a constant- γ ideal-gas approximation. Each component is solved with these variable- C_p properties: an isentropic outlet state is found by solving the exact entropy balance for temperature (root-found, no constant- γ shortcut), and the actual outlet state follows from an isentropic-efficiency correction to the ideal specific work. For the compressor ($2 \rightarrow 3$), given pressure ratio Π_c and isentropic efficiency η_c from the map lookup of Section 2.1,

$$\Delta h_{is,c} = h(T_{3s}) - h(T_2) \quad (1)$$

$$\Delta h_{act,c} = \frac{\Delta h_{is,c}}{\eta_c} \quad (2)$$

$$h_3 = h_2 + \Delta h_{act,c} \quad (3)$$

with $T_3 = h^{-1}(h_3)$ and shaft power absorbed $\dot{W}_c = \dot{m}\Delta h_{act,c}$. The turbine ($4 \rightarrow 5$) mirrors this, but the efficiency multiplies rather than divides the ideal work, since a real turbine extracts less than the ideal machine:

$$\Delta h_{is,t} = h(T_4) - h(T_{5s}) \quad (4)$$

$$\Delta h_{act,t} = \eta_t \Delta h_{is,t} \quad (5)$$

$$h_5 = h_4 - \Delta h_{act,t} \quad (6)$$

Hydrogen is burned across the combustor ($35 \rightarrow 4$) to raise the recuperator-preheated air to the specified firing temperature (TIT); the steady-flow energy balance per unit air mass, with fuel/air ratio FAR and combustion efficiency η_{comb} ,

$$h_{35} + FAR \cdot (\eta_{comb} LHV_{H_2}) = (1 + FAR)h_4(TIT) \quad (7)$$

is iterated to self-consistency because h_4 depends on the burned-gas composition, which itself depends on FAR. The recuperator ($3, 5 \rightarrow 35, 6$) is modelled as a counter-flow exchanger with effectiveness ε relative to the thermodynamic maximum,

$$C_{min} = \min(\dot{m}_c C_{p,c}, \dot{m}_h C_{p,h})|_{\bar{T}} \quad (8)$$

$$\dot{Q} = \varepsilon C_{min} \max(T_5 - T_3, 0) \quad (9)$$

with outlet states from energy balances on each side, and since T_{35} depends on T_5 (equation (9)), T_5 on the turbine expansion from TIT (equation (6)), and the fuel flow needed to reach TIT on T_{35} (equation (7)), the recuperator/combustor/turbine block is closed by fixed-point iteration (tolerance 0.01 K). Overall shaft, electrical, and thermal-efficiency figures follow as

$$\dot{W}_{shaft} = (\dot{W}_t - \dot{W}_c)\eta_{mech} \quad (10)$$

$$\dot{W}_{elec} = \dot{W}_{shaft}\eta_{gen} \quad (11)$$

$$\eta_{th} = \frac{\dot{W}_{elec}}{\dot{m}_{fuel} LHV_{H_2}} \quad (12)$$

with fixed cycle constants $\pi_{inlet} = 0.997$, $\eta_{gen} = 0.95$, $\eta_{mech} = 0.99$, cold/hot recuperator pressure losses of 2.5%/3.0%, combustion efficiency 0.999, combustor pressure loss 3.0%, and $LHV_{H_2} = 120$ MJ/kg [6, 11].

2.1. Characteristic maps and calibration

Component efficiency and flow are not treated as constants: they are read from manufacturer-style compressor and turbine performance maps, each a beta-line-format lookup table on an 11×16 grid. Corrected compressor speed uses a variable- γ similarity correction and the operating point is located by a one-dimensional root-find on the map's beta-line coordinate; on the turbine side both map axes are already relative, so the operating point is located directly from a gamma-corrected isentropic work-function ratio and a corrected-speed term. Because a generic reference map does not exactly represent one specific unit, ten scale factors (five uniform corrections and five terms varying linearly, or for the recuperator linearly plus quadratically, with corrected speed) adapt the maps and the recuperator effectiveness to the AE-T100's own measured behaviour. These ten free parameters are fitted by a multi-start Nelder–Mead simplex search minimizing a weighted sum of relative-% RMSEs across compressor outlet temperature (COT), recuperator cold outlet/combustor inlet temperature (CCIT), turbine outlet temperature (TOT), and electrical power, against six measured steady-state operating points (50–100

kW) from a test campaign, with weights 1.0/2.0/2.0/1.5 respectively (CCIT and TOT weighted highest as the historically worst-fitting channels). Table 2 gives the locked, calibrated values used for all subsequent work.

Table 2. Calibrated map-scaling parameters.

Parameter	Calibrated value
Compressor efficiency scale, a_{η_c}	0.9798
Turbine efficiency scale, a_{η_t}	0.9693
Compressor flow scale, $a_{\dot{m}}$	0.9707
Compressor pressure-ratio scale, a_{Π}	1.0504
Recuperator effectiveness (design speed), ε_0	0.8833
Flow speed-slope, $b_{\dot{m}}$	0.8340
Compressor efficiency speed-slope, b_{η_c}	0.0320
Turbine efficiency speed-slope, b_{η_t}	-0.1160
Recuperator linear/quadratic speed-slopes, $b_{\varepsilon,1}, b_{\varepsilon,2}$	-0.4013, -1.5462

2.2. Calibration agreement with test-rig data

The six measured operating points (50–100 kW) are the only steady-state data available, so they serve both to fit the ten parameters and to assess the result; the resulting errors are therefore reported as calibration agreement, not as independent validation. For each point, the measured ambient conditions, compressor pressure ratio, shaft speed and TIT are fed in as inputs, and COT, CCIT, TOT and electrical power are predicted. TIT is not measured on the AE-T100 (figure 1): the values in the test-campaign record are those calculated by the unit's control system from measured quantities, chiefly TOT, and they are used only as a model input, never as a comparison target. Table 3 gives the resulting root-mean-square errors and figure 2 shows the calibrated trends across the load range. Because the ten parameters were fitted to these same six points, these errors show how closely the model reproduces the measurements; they are not an estimate of its accuracy at unmeasured operating points. All six points were recorded at compressor inlet temperatures of 8–13 °C, so model accuracy at other ambient temperatures has not been checked against measurement (section 6).

Table 3. Calibration agreement (RMSE across the six measured points).

Channel	RMSE
COT (compressor outlet temperature)	0.32 K
CCIT (recuperator cold outlet / combustor inlet)	1.20 K
TOT (turbine outlet temperature)	0.94 K
Electrical power	0.87%

AE-T100 MGT: Model vs Measured Data

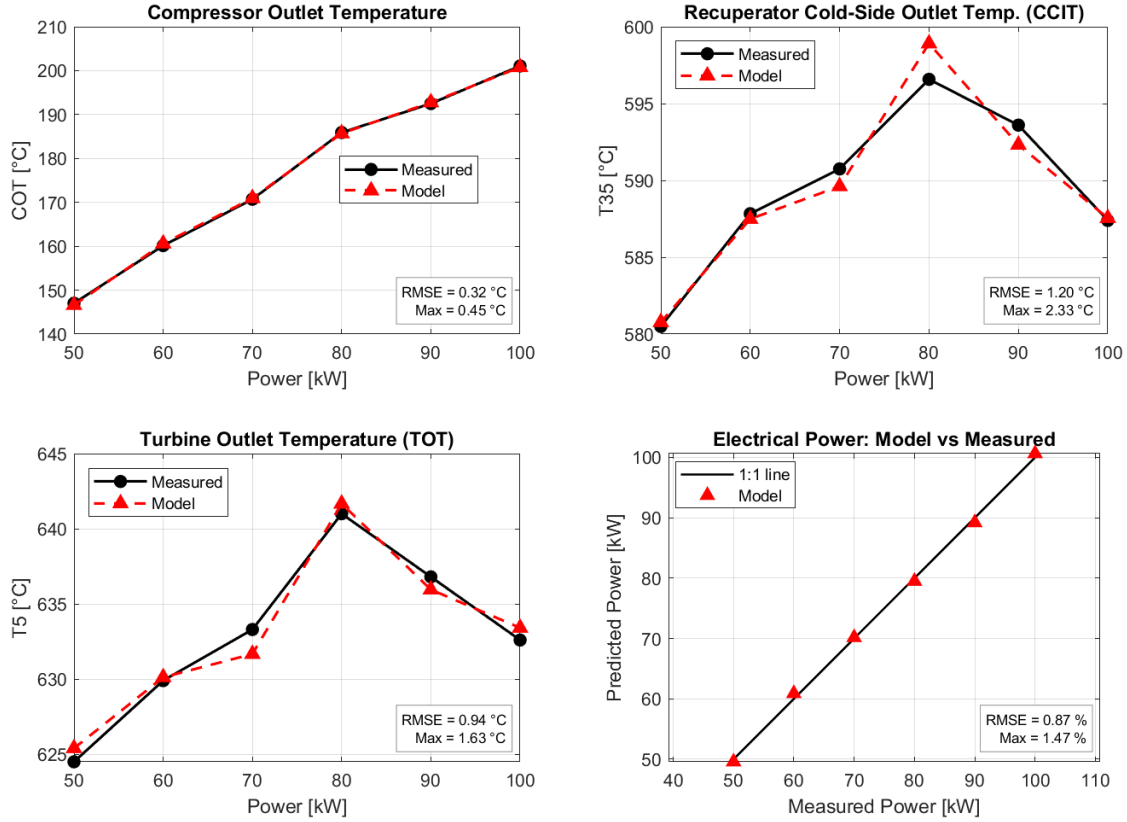


Figure 2. Measured vs. calibrated-model COT, CCIT, TOT, and electrical power across the six operating points, with in-sample RMSE and worst-case error annotated on each panel.

3. Degradation modelling

Thermal degradation is injected into the healthy model through a multiplicative degradation vector applied inside the component functions of Section 2. Three fault classes are defined, each as a primary degradation with a jittered, coupled secondary degradation: compressor fouling (CF, primary flow-capacity loss $|\delta\dot{m}_c|$ of 0.5–6%, coupled to an efficiency loss $\delta\eta_c = -r|\delta\dot{m}_c|$ with $r \in [0.5, 1.0]$), turbine erosion (TE, primary efficiency loss $|\delta\eta_t|$ of 0.5–6%, coupled to a swallowing-capacity increase $\delta\dot{m}_t = +r|\delta\eta_t|$ with $r \in [0.3, 0.7]$), and recuperator degradation (RD, primary effectiveness loss $|\delta\epsilon|$ of 0.5–9%, coupled to a pressure-loss increase with $r \in [0.2, 0.5]$). Severity is normalized to $[0, 1]$ as the primary delta magnitude divided by the class maximum. Every run, including Healthy, additionally carries an independent $\pm 0.5\%$ uniform jitter on all deltas simultaneously, representing unit-to-unit manufacturing and wear scatter; the 0.5% severity floor on fault classes sits just above this jitter band, so the Healthy/faulted boundary is well-defined by construction rather than by an arbitrary threshold.

The turbine swallowing-capacity delta $\delta\dot{m}_t$ requires care: at a matched steady operating point the turbine must pass exactly the mass flow the combustor delivers, so a capacity fault cannot simply create an internal flow discontinuity. Instead it is applied as a multiplicative correction to the turbine's swallowing capacity within the flow-matching constraint of the prediction-mode solver (Section 4), shifting the whole engine's solved operating point (speed, TIT, compressor beta) to a new equilibrium – the behaviour a real capacity-limiting fault (e.g. partial blade damage) would produce.

4. Synthetic fault dataset generation

Generating fault data requires the model to close the loop itself, emulating the T100's own controller: given only ambient conditions and a commanded electrical power setpoint, three unknowns (shaft speed,

TIT, and the compressor map beta coordinate) are solved by a damped Newton iteration against three constraints – electrical power matches the setpoint, turbine outlet temperature follows the controller’s measured TOT-vs-power schedule, and the turbine’s swallowing capacity matches the delivered flow (the capacity itself read from the turbine map and anchored to the six measured points). This prediction-mode solver was confirmed to recover the measured healthy operating line (shaft speed within ± 0.2 percentage points, pressure ratio within ± 0.02 , TIT within ± 2.5 K of the value calculated by the control system) before being used to generate fault data.

The sampling grid spans the measured load range (50–100 kW) but extends the ambient temperature to 0–40 °C, well beyond the 8–13 °C of the calibration data, so ambient conditions outside that band rely on extrapolation of the calibrated model: 6 loads $\times$ 5 ambient temperatures = 30 operating conditions, $\times$ 4 classes (Healthy, CF, TE, RD) $\times$ 50 draws = 6,000 planned physical solves. Of these, 1,410 did not yield a valid steady state, for two different reasons. Only one, a severe TE run at 40 °C/50 kW, was a numerical non-convergence: the Newton iteration stopped without satisfying its constraints while all unknowns stayed inside their bounds. The other 1,409 reached the shaft-speed limit ($1.01 \times 70,000$ rpm) before delivering the demanded power, a physically infeasible operating point at which a real machine would derate. These runs are concentrated in hot, high-load conditions and are not class-balanced (TE 496, CF 436, RD 240, Healthy 238). Exclusion rises with severity for the turbomachinery faults, from 21% to 39% of CF runs and from 23% to 44% of TE runs between the lowest and highest severity thirds, while RD stays near 16%; at 20 °C and below, at most 10% of any class is lost. As a result, valid high-severity runs exist at only 67% (CF), 63% (TE) and 87% (RD) of the 30 conditions, so severely degraded machines in hot, high-load operation, the state in which a real unit would derate, are under-represented. The 4,590 valid runs are each expanded into 5 independently-seeded noisy realizations using an additive Gaussian model for temperatures (1.5 K for CIT/COT; 2.5 K, a type-K thermocouple-class value, for CCIT/TOT/EGT) [4, 5] and a multiplicative Gaussian model for pressures (0.3%), electrical power (0.5%), shaft speed (0.05%), and fuel flow (1%).

The final labelled dataset contains $4,590 \times 5 = 22,950$ rows: Healthy 6,310, CF 5,320, RD 6,300, TE 5,020. The machine-learning feature set is the eleven noisy sensor channels a deployed unit would actually carry – compressor inlet temperature and pressure, shaft speed, electrical power, compressor outlet temperature and pressure, recuperator cold-outlet temperature and pressure, turbine outlet temperature, exhaust gas temperature, and fuel mass flow – plus the known commanded power setpoint; injected degradation deltas, solved states, and healthy-reference component efficiencies are retained only as oracle/diagnostic columns and excluded from classifier input. Fault-signature directions were checked against expectations from classical gas-path analysis: compressor fouling raises shaft speed and COT while lowering flow; turbine erosion raises speed even more and lowers turbine efficiency; recuperator degradation lowers CCIT and raises EGT as less heat is recovered, with fuel flow rising modestly across all three classes to hold the controller’s power/TOT targets.

5. Random forest fault classifier

The classifier is implemented in Python using scikit-learn. Its input is restricted to what a deployed unit could observe: the eleven noisy sensor channels of Section 4 plus the known commanded power setpoint. From these twelve, eleven physics-motivated quantities are derived – compressor pressure ratio and temperature rise, an isentropic-efficiency proxy $\frac{T_{3s}-T_2}{T_3-T_2}$, a recuperator-effectiveness proxy and a hot-side temperature-drop ratio, the cold-side pressure-loss fraction, corrected shaft speed, specific fuel consumption, an electrical-efficiency proxy, the turbine-to-exhaust temperature drop, and the controller power error – giving 23 base features. Each base feature is then paired with its residual against the healthy mean of the same operating condition, which removes operating-point variation and isolates the fault signature, for 46 features in total. Those healthy reference means are fitted inside the model pipeline on training healthy rows only, so no validation or test information reaches the baseline at any point, including within cross-validation folds.

In this configuration the healthy reference is a lookup table of per-condition means, defined only at the 30 grid conditions; an unseen condition falls back to the global healthy mean. For use at arbitrary conditions, a second, interpolated scheme was also tested, in which each healthy reference is a quadratic surface in the two measured context variables, compressor inlet temperature and power setpoint, fitted by ridge regression to the training healthy rows. Each per-condition reference is the mean of up to 175 healthy training rows (about 160 on average), so its noise-induced uncertainty is roughly $\sigma/\sqrt{n}$, about 0.2 K for the 2.5 K thermocouple channels and small compared with the fault signatures. Gradual healthy deterioration is represented only by the $\pm 0.5\%$ jitter band of the Healthy class; in service the reference would have to be re-established after each overhaul or compressor wash and tracked over time, since slow wear would otherwise gradually be reported as a fault.

The train/test split is a 70/30 grouped shuffle split by physical run rather than by row – 16,065 training rows (3,213 runs) against 6,885 held-out test rows (1,377 runs) – since a naive row-level split would place noise-realization siblings of the same operating point on both sides and leak information. Hyperparameters (maximum features per split, minimum samples per leaf, maximum tree depth, and class weighting) were selected by a 20-candidate randomized search over 400-tree forests under 5-fold grouped cross-validation on the training runs alone. The selected configuration was then refitted as a 1,000-tree forest and evaluated once on the held-out test runs.

The model reaches 90.1% test accuracy with a macro-F1 of 0.896. Table 4 and Figure 3 give its confusion matrix, and Table 5 its per-class precision, recall, and F1. Recuperator degradation is the most cleanly identified class (precision 0.977, recall 0.945), and the two turbomachinery classes are the hardest, confusing chiefly with each other and with Healthy – expected, since a compressor flow-capacity loss and a turbine efficiency loss both push the operating point in broadly the same direction (higher shaft speed and higher fuel flow at fixed power and TOT). Collapsing to a binary fault/healthy decision, the classifier detects 93.7% of faulty rows at a 5.9% false-alarm rate. Figure 4 ranks the features by mean decrease in impurity. The recuperator cold-outlet temperature residual and the shaft-speed residual dominate, followed by the engineered hot-side temperature-drop ratio and the compressor isentropic-efficiency proxy, while neither turbine outlet temperature nor electrical power reaches the top fifteen – consistent with the physics, since the controller regulates both of those to their targets regardless of engine health.

Table 4. Confusion matrix, RF test set (rows = true, columns = predicted).

True \ Pred.	Healthy	CF	TE	RD
Healthy	1826	70	17	27
CF	154	1220	90	11
TE	67	128	1238	7
RD	93	14	4	1919

Table 5. Per-class precision, recall, and F1 (RF test set).

Class	Precision	Recall	F1
Healthy	0.853	0.941	0.895
CF	0.852	0.827	0.839
TE	0.918	0.860	0.888
RD	0.977	0.945	0.961

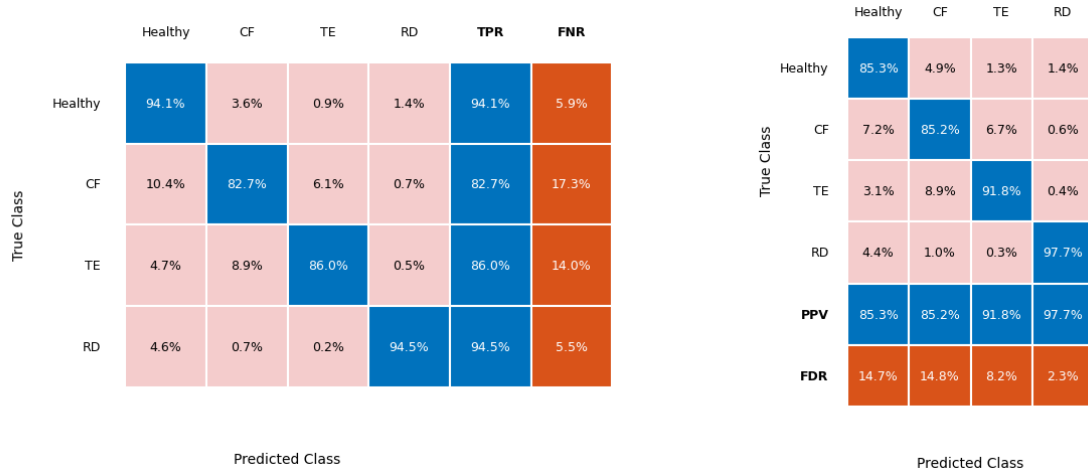


Figure 3. RF test-set confusion matrices. Left: row-normalized, with true-positive and false-negative rates per true class. Right: column-normalized, with positive predictive value and false discovery rate per predicted class.

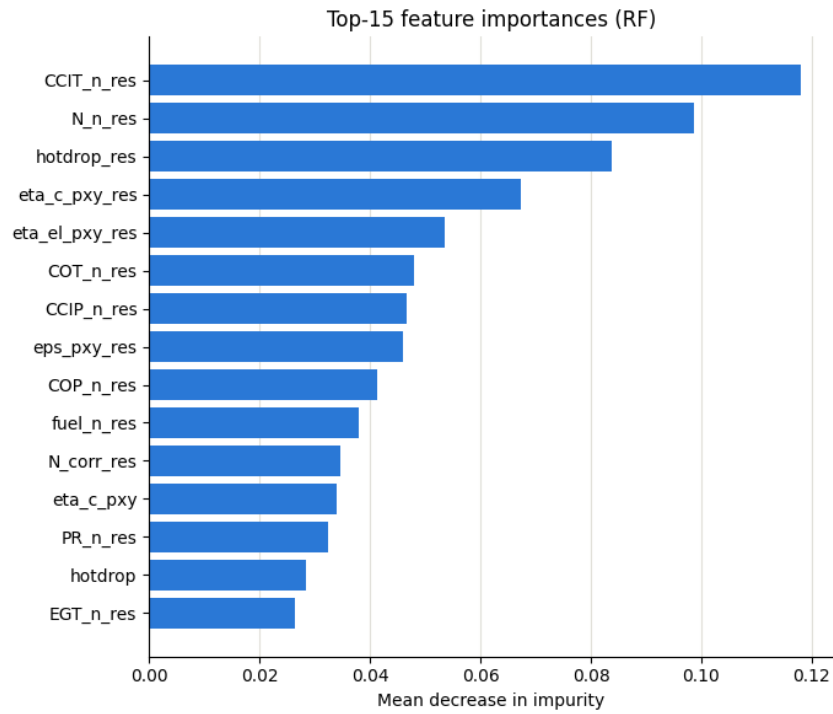


Figure 4. Top-15 random-forest feature importances by mean decrease in impurity; res denotes a healthy-referenced residual feature.

Uncertainty was estimated at the level of physical runs rather than rows. On a re-run of the pipeline (89.9% accuracy, within the interval below), resampling the 1,377 test runs with replacement (2,000 bootstrap draws) gives 95% confidence intervals of 88.6–91.1% for accuracy, 0.881–0.906 for macro-F1, 92.0–94.6% for fault detection and 4.3–7.5% for false alarms, and ten further run-grouped splits with different seeds give $89.0 \pm 0.4\%$ accuracy. Averaging class probabilities over the five noise realizations of each run raises accuracy to 93.3%. Table 6 compares the forest with simpler classifiers on the same features and split, and with reduced feature sets. A linear logistic regression slightly exceeds the forest in accuracy (91.1%) but at a higher false-alarm rate (8.2% against 5.9%), gradient boosting performs alike, and LDA, k-nearest neighbours and a single tree reach 84–85%. The ablation shows

where the accuracy comes from: raw sensors alone give 82.2%, the engineered physics features raise this to 85.8%, and the healthy-referenced residuals bring it to 89–90%. Most of the diagnostic information is therefore created by the physics-based residual formulation, not by the choice of classifier; the forest is retained for its lower false-alarm rate. Removing every fuel-flow-based feature costs only 0.3 percentage points.

Table 6. Benchmark classifiers and feature ablation (same 46-feature pipeline and run-grouped split unless stated).

Configuration	Accuracy	Macro-F1	Detection	False alarm
Majority class (always Healthy)	28.2%	0.110	0.0%	0.0%
LDA	83.6%	0.842	79.5%	1.8%
k-NN (k=15)	84.2%	0.836	86.2%	4.9%
Decision tree	84.8%	0.841	93.9%	17.9%
Logistic regression	91.1%	0.907	94.9%	8.2%
Gradient boosting	89.8%	0.894	94.3%	8.7%
Random forest	89.9%	0.894	93.3%	5.9%
RF, raw sensors + setpoint (12)	82.2%	0.805	91.5%	7.0%
RF, raw + engineered (23)	85.8%	0.849	91.6%	8.6%
RF, sensors + sensor residuals (24)	89.0%	0.883	92.8%	5.4%
RF, residuals only (23)	89.2%	0.887	93.3%	9.0%
RF, full set without fuel-flow features (40)	89.6%	0.892	93.1%	6.6%

Performance depends mainly on fault severity. In the lowest severity third, detection is 66% for CF, 85% for TE and 83% for RD, and 53%, 62% and 81% of those cases respectively receive the correct class; above one third of the maximum severity, detection is at least 98% and correct classification at least 92% for every class. Accuracy varies much less with operating condition, between 88.6% and 91.3% across the five ambient temperatures and six load levels.

The run-grouped split keeps noise realizations together but still shares operating conditions between training and test. Generalization to unseen conditions was therefore tested by holding out complete load–temperature combinations (table 7). With the lookup baseline, accuracy collapses to about 34%, with about 90% of healthy rows flagged as faulty, because an unseen condition falls back to the global healthy mean and the residuals then reflect the change of operating point rather than degradation. With the interpolated baseline, accuracy is 81.5% when single conditions are withheld, 79.3% when a whole load level is withheld, and 76.5% when a whole ambient temperature is withheld, the last including

extrapolation to 0 °C and 40 °C. The interpolated baseline costs accuracy at seen conditions (86.0% on the random split), so the lookup suits a machine that runs at a few fixed conditions and interpolation one that does not. To test whether unequal class availability explains part of the performance, a forest was given only the operating condition (inlet temperature and setpoint): it scores 26.7%, below the 29.5% share of the most frequent test class (RD), so the condition alone carries no usable class information. The main model scores 90.6% on the 14 conditions where all four classes are complete and 88.7% on the 13 others, and a model retrained on the complete conditions alone scores 90.4% on them; unequal availability therefore does not inflate the headline result.

Table 7. Generalization to operating conditions withheld from training (random forest).

Test	Healthy baseline	Accuracy	Macro-F1	False alarm
Random run-grouped split	Lookup	89.9%	0.894	5.9%
Random run-grouped split	Interpolated	86.0%	0.852	7.9%
Leave one condition out	Lookup	33.8%	0.311	90.0%
Leave one condition out	Interpolated	81.5%	0.813	15.8%
Leave one load out	Lookup	33.8%	0.316	86.0%
Leave one load out	Interpolated	79.3%	0.790	19.5%
Leave one ambient out	Lookup	33.2%	0.302	91.9%
Leave one ambient out	Interpolated	76.5%	0.764	21.6%

6. Discussion and limitations

- TOT schedule assumed ambient-independent.** The turbine-outlet-temperature schedule used to close the prediction-mode solver is interpolated from six campaign points measured at 8–13 °C ambient and applied unchanged across the full 0–40 °C generation range. No measured TOT-vs-ambient data exists to confirm this directly; it is consistent with the T100’s known constant-TOT control philosophy, and the regime where it would matter most (hot, high-load operation) is exactly the regime the solver already excludes via the speed-cap derating. **More generally, the calibration data cover 8–13 °C ambient only while the synthetic data span 0–40 °C, so all results outside that band rest on extrapolation of the calibrated model and have not been checked against measurement.**
- Hot-corner class-coverage shortcut risk.** Because degraded engines hit the shaft-speed solver bound earlier than healthy ones, class availability correlates with operating condition in hot, high-load corners, **and three conditions (30 °C/100 kW and 40 °C/90–100 kW) have no samples of any class; the group-level train/test split does not fully control for this, since it splits by run rather than by condition.**
- Steady-state, single-fault, literature-typical scope.** The dataset contains only steady-state, single-fault operating points – no transients, simultaneous faults, or sensor faults (drift, bias,

stuck readings) – and the fault severity ranges and coupling ratios were chosen to be physically plausible and directionally consistent with classical gas-path analysis, not fit to measured degraded-unit data, since none exists for this machine. Model fidelity itself rests on a [six-point calibration envelope](#) at near-ambient conditions, so extrapolation to the full 0–40 °C range inherits whatever model error exists [outside that calibrated envelope](#).

- **Noise model, sensor bias and model error.** Independent Gaussian noise does not represent sensor bias, drift or thermodynamic-model error. A test-time bias check (model trained on unbiased data, one channel offset at a time) shows the classifier is insensitive to TOT and power offsets, which the controller holds at their targets, but sensitive to the compressor and recuperator temperatures: a +3 K COT bias lowers accuracy to 83.5% and raises false alarms to 16.7%, and a +5 K EGT or –5 K CCIT bias raises false alarms to 19–22%. Drift acts as a slowly growing bias of the same kind. Model error was not simulated; it acts as a systematic bias on all channels at once, mainly shifting the healthy reference, which should therefore be re-anchored on measured healthy data from the target unit before deployment. Training with randomized bias and explicit sensor-fault classes is left to future work.
- **Closed-set classification.** The classifier is trained and evaluated on exactly four classes and has no mechanism to flag or reject an input that does not resemble any trained class; a fault mode outside these three (e.g. bearing damage, sensor failure, a combined fault) would simply be forced into the nearest trained label rather than recognized as novel. An explicit open-set/novelty-rejection stage ahead of the four-way classifier is a natural next step before any deployment context where untrained fault modes are plausible.

7. Conclusion

A physics-based thermodynamic model of a hydrogen-fuelled AE-T100 micro gas turbine was calibrated to six measured operating points, with calibration agreement of 0.32–1.20 K on gas-path temperatures and 0.87% on electrical power, and was used to generate a 22,950-row labelled fault dataset spanning four health classes. Because the measurements cover 8–13 °C ambient only, the 0–40 °C dataset relies on extrapolation of the calibrated model. A random forest trained on eleven realistic sensor channels and evaluated under a leakage-safe, run-grouped split reached 90.1% test accuracy (macro-F1 0.896), with a 93.7% fault-detection rate at a 5.9% false-alarm rate; most of this accuracy comes from the healthy-referenced residual features, and a logistic regression on the same features performs similarly. With complete operating conditions withheld from training, accuracy was 77–82% using an interpolated healthy baseline, with false-alarm rates rising to 16–22%. The approach offers a route to labelled fault data and diagnostic models for machines where deliberately degraded operating data do not exist, but field use requires validation against real degradation and robustness to sensor bias and model error. Future work includes an open-set/novelty-rejection stage, combined faults, combustor faults and hydrogen–natural-gas blends, and additional measured operating points, especially at other ambient temperatures.

Acknowledgments

This work was supported by the AIESI project (NORPART-2021/10403), funded by the Norwegian Directorate for Higher Education and Skills (HK-dir) under the NORPART scheme. The views expressed are those of the authors alone, and HK-dir bears no responsibility for any use made of the information presented here.

References

- [1] Kurz R and Brun K 2001 Degradation in gas turbine systems *J. Eng. Gas Turbines Power* 123 70–7
- [2] Meher-Homji C B and Bromley A 2004 Gas turbine axial compressor fouling and washing *Proc. 33rd Turbomachinery Symp.* (Texas A&M Turbomachinery Laboratories) 163–92
- [3] Talebi S S, Madadi A and Tousi A M 2025 Fault diagnosis of micro gas turbine using a hybrid

- scheme of thermodynamic analysis and artificial neural network *Heliyon* 11 e43378
- [4] Hashmi M B, Mansouri M, Fentaye A D and Ahsan S 2023 A deep learning approach for fault diagnosis of hydrogen fueled micro gas turbines *Proc. SIMS 64* (Västerås, Sweden, 26–27 September 2023)
 - [5] Hashmi M B, Mansouri M, Fentaye A D, Ahsan S and Kyprianidis K 2024 An artificial neural network-based fault diagnostics approach for hydrogen-fueled micro gas turbines *Energies* 17 719
 - [6] Banihabib R and Assadi M 2022 A hydrogen-fueled micro gas turbine unit for carbon-free heat and power generation *Sustainability* 14 13305
 - [7] McBride B J, Zehe M J and Gordon S 2002 NASA Glenn coefficients for calculating thermodynamic properties of individual species NASA/TP-2002-211556
 - [8] Turbec AB 2003 T100 microturbine CHP system: technical description D10293 Ver 4.0
 - [9] Fentaye A D, Baheta A T, Gilani S I and Kyprianidis K G 2019 A review on gas turbine gas-path diagnostics: state-of-the-art methods, challenges and opportunities *Aerospace* 6 83
 - [10] Xiao G, Yang T, Liu H, Ni D, Ferrari M L, Li M, Luo Z, Cen K and Ni M 2017 Recuperators for micro gas turbines: a review *Appl. Energy* 197 83–99
 - [11] Alhuyi Nazari M, Fahim Alavi M, Salem M and El Haj Assad M 2022 Utilization of hydrogen in gas turbines: a comprehensive review *Int. J. Low-Carbon Technol.* 17 513–9
 - [12] Molla Salilew W, Abdul Karim Z A and Alemu Lemma T 2022 Investigation of fault detection and isolation accuracy of different machine learning techniques with different data processing methods for gas turbine *Alexandria Eng. J.* 61 12635–51