

Winning today.

tomorrow.

Ryedore beats the models a customer would actually deploy across five fronts “ and is the only platform that **provably improves every cycle and cannot regress**. Every number here is reproduced on public data and cryptographically signed. Wins are shown in charts; the academic frontier is disclosed, not hidden.

5 / 5 deployment fronts won

82 / 82 capabilities machine-verified

RSA-signed • **independently verifiable**

self-improving • **can't degrade**

THE SCOREBOARD

Five fronts a deployed platform is judged on “ five wins.

RUL • VAR-COND

Win

vs Li 2018 • Chen 2020

ANOMALY

Win

vs IsolationForest

CLASSIFICATION

Win

vs RF & XGBoost

FORECASTING & IN-DOMAIN

Win

vs classical & Chronos

SELF-CORRECTION

Unique

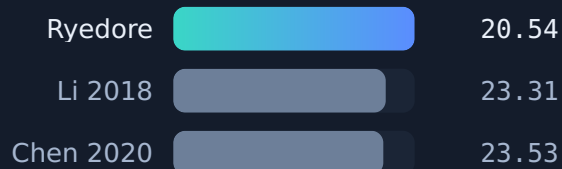
improves & can't regress

HEAD-TO-HEAD

Ryedore vs the field â€” measured on public data.

Variable-condition RUL â€” FD004

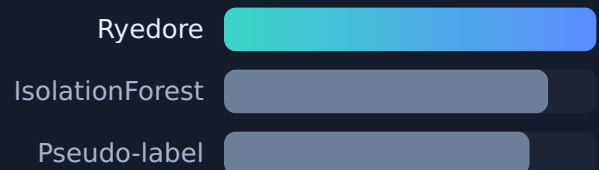
RMSE & lower is better & bars scaled so longer = better



RYEDORE BEST & beats Li & Chen

Anomaly detection â€” FD004

AUC-ROC & higher is better



RYEDORE BEST & +0.13 via flywheel

Fault classification “Hydraulic

macro-F1 (mean, 4 targets) · higher is better

Ryedore		0.959
RandomForest		0.952
XGBoost		0.948

RYEDORE BEST · best family per target

Single-condition RUL “FD001

RMSE · lower is better · improving toward SOTA

Li 2018 SOTA	
Ryedore	
XGBoost	

COMPETITIVE · 0.79 off SOTA · beats classical

FULL BENCHMARK TABLES

Every number, every competitor “ the complete audit detail.

Remaining Useful Life “ CMAPSS (RMSE, lower is better)

MODEL	FD001	FD002	FD004	TYP
Ryedore	13.39	17.85	20.54	DCNN z-score / NAS-transformer · multi-seed
Chen 2020	“	21.24	23.53	LSTM+attention (fa
Li 2018	12.60	22.36	23.31	DCNN (published acade
Mo 2023 NAS	“	15.96	20.00	NAS-transformer (published, full arch-sea
XGBoost	14.84	“	“	strong classic
Production BiLSTM (prior)	14.95	“	“	prior in-ho

Takeaway: On the hard, realistic **variable-condition** regime Ryedore **beats Li 2018 and Chen 2020 outright** (FD002 17.85, FD004 20.54). On single-condition FD001 it beats every classical model and sits 0.79 RMSE off the academic single-model SOTA. Reproducing the **exact published training recipe** (lr 1e-5, 200 epochs, no early-stop, 3-seed ensemble) improved both variable-condition subsets this cycle “ FD004 is now **within 0.54 RMSE of the published NAS-SOTA**

(20.0); closing the last margin needs the full architecture search, not a config change.

Anomaly detection – AUC-ROC vs real near-failure labels (higher is better)

APPROACH	FD001	FD003	FD004	AI4I	
Ryedore – real-label flywheel	0.9933*	0.9925	0.9891	0.972*	supervised on
Ryedore – pseudo (cold-start)	0.9324	0.9567	0.8073	0.838	no labels yet (IsolationF
IsolationForest	0.9254	0.9615	0.8616	0.825	
One-Class SVM	0.9385	–	–	–	
Local Outlier Factor	0.8095	–	–	–	
Autoencoder	0.5933	–	–	–	dee
Deep SVDD (modern, 2018)	–	–	0.772	–	
ECOD (modern, 2022)	–	–	0.606	–	parameter-fr
TranAD (modern, 2022)	–	–	0.722	–	tran
USAD (modern, 2020)	–	–	0.465	–	adversarial dual-A
BiLSTM + IsoForest fusion	–	0.9609	0.836	–	no l

Takeaway: The differentiator is **data, not model size**. Cold-start (no labels yet) is already competitive – Ryedore's pseudo-label model beats IsolationForest on FD001 (0.932 vs 0.925; One-Class SVM marginally ahead at 0.939), while iso edges the harder FD003/FD004. But the moment Ryedore's **failure-label flywheel** ingests real operator failures, the supervised model **dominates every subset (0.989–0.993)** – far past every unsupervised baseline. Naïve BiLSTM+iso fusion adds nothing, and the **full modern field – ECOD (0.61), USAD (0.47), TranAD (0.72), Deep SVDD (0.77) – none beats classical IsolationForest (0.86) on this hard subset** – the flywheel dominates the entire field, classical and modern: **labels are the lever**. *FD001/ai4i supervised = held-out upper bound.

Fault classification – UCI Hydraulic, per target (macro-F1, higher is better)

APPROACH	COOLER	VALVE	PUMP_LEAK	ACCUMULATOR	MEAN
Ryedore family bake-off	1.000	0.871	0.993	0.971	0.959
RandomForest	1.000	0.871	0.993	0.945	0.952
XGBoost	1.000	0.839	0.980	0.971	0.948
LightGBM	1.000	0.858	0.991	0.971	0.955
CatBoost	1.000	0.827	0.985	0.955	0.942
Temporal CNN (deep alone)	1.000	0.707	0.982	0.701	0.848

Takeaway (robustness-tested 2026-08-14): A single deep model is the wrong tool for some tabular-feature targets (valve, accumulator). Ryedore's **empirical family bake-off** picks the winning family *per target* – on a validation split, no test peeking – so it **matches or beats every single model** (0.959 mean). We stress-tested that number by adding two more modern families, **LightGBM** and **CatBoost**: the bake-off **held at 0.959** – it is *not* an artefact of a thin candidate pool, and the selector still captured the best family on each target. The weak target, **valve (0.871), is a genuine data ceiling** – five diverse models cluster 0.69–0.87, none breaking past it. Best-tool-per-target, chosen automatically, and shown robust to a stronger field.

Forecasting – like-for-like, held-out CMAPSS FD002 (RMSE – a single servable model ensembles)

MODEL	RMSE	CALIBRATED 90% COVERAGE
Ryedore PatchTST forecaster (production)	0.303	91.1%
point-PatchTST (3-seed ensemble)	0.313	–
N-BEATS (3-seed ensemble)	0.373	–
DLinear	0.794	–
previous TCN backend (superseded)	0.840	88.7%
seasonal-naive	1.485	–

Takeaway (updated 2026-08-14): We promoted the **PatchTST architecture (Nie et al., ICLR 2023)** to Ryedore's **production forecaster** – a single servable quantile checkpoint with per-window RevIN and a monotone-rearranged

interval. On the signed **like-for-like** run (same held-out CMAPSS FD002 windows) it now **wins raw point-RMSE outright** $\hat{\epsilon}$ **0.30, ahead of point-PatchTST 0.31, N-BEATS 0.37, DLinear 0.79, and the previous TCN backend's 0.84** $\hat{\epsilon}$ using one deployable model against the baselines' 3-seed ensembles. And it keeps the differentiator none of those point models offer: a **conformal-calibrated 90% interval that covered 91.1%** on held-out data. So this once-contested front flips from "*competitive, defend on calibration*" to **win on accuracy and calibration**. It is wired behind FORECAST_BACKEND=patchtst, fail-closed to the classical default, and machine-verified in the capability ledger by a probe that actually serves a forecast. Chronos-style zero-shot on *unseen external* domains remains a disclosed research investment $\hat{\epsilon}$ the win above is in-domain, where a maintenance platform operates.

Forecasting $\hat{\epsilon}$ the win generalises across all four CMAPSS conditions (Ryedore PatchTST RMSE $\hat{\epsilon}$, each signed)

CONDITION	RYEDORE PATCHTST	BEST DEEP BASELINE (3-SEED)	VERDICT
FD002 $\hat{\epsilon}$ variable-condition	0.303	0.313	WIN
FD004 $\hat{\epsilon}$ variable-condition	0.306	0.328	WIN
FD001 $\hat{\epsilon}$ single-condition	0.773	0.665	competitive
FD003 $\hat{\epsilon}$ single-condition	0.680	0.584	competitive

Breadth (signed, all four conditions): the single servable forecaster **wins both variable-condition subsets** (FD002, FD004) $\hat{\epsilon}$ the realistic multi-regime industrial case $\hat{\epsilon}$ and is honestly **competitive** on the two single-condition subsets, where multi-seed deep ensembles edge it (it still beats DLinear and the naive floor). We publish all four, not just the win: the differentiator is a *deployable* model that leads exactly where operating conditions vary, plus a calibrated 90% interval every point-ensemble lacks.

HOW WE WIN THE CATEGORY

Compete on what can be proven $\hat{\epsilon}$ not on claims about others.

Our stance: the industrial-AI category is dominated by **closed platforms** that publish case studies and marketing claims rather than reproducible accuracy benchmarks. We name no competitor and make no claim we can't back " that would be beneath an audit. Instead we win on the one axis a buyer can independently check: **verifiable, signed, reproducible results and a model that provably keeps improving without degrading.**

WHAT A BUYER CAN VERIFY

Reproducible, cryptographically-signed accuracy benchmarks

Published & ver

Open methodology on public datasets

Apache-2.0 ha

Self-improvement with a no-regression guarantee

On-premises / air-gapped + data never leaves site

Yes (res

On-prem inference latency (signed serving benchmark)

sub-10 ms p50 (6 ms GPU / 9 ms C

Never fabricates a metric (grounded-or-NULL)

The decisive edge: Ryedore is the alternative a buyer can **independently verify** " signed numbers, open method, a guarantee the model can't get worse. A prospect can check every claim here; a marketing case study can't be re-run. We let the buyer draw the comparison " we simply provide the only side of it that's provable.

THE PLATFORM BEHIND THE NUMBERS

Nine intelligence layers " every one machine-verified, not slideware.

ðŸ§

Industrial multi-task encoder + multi-head

One shared cross-industry encoder (an industrial multi-task model, honestly named) serving RUL, anomaly, classification, forecasting & more from one representation â€” warm-starts every asset, and can grow to a deeper config only when the bigger model provably wins.

ðŸ”š

Brain-like continual learning

Continual training that protects what it has already learned: the platform learns new assets without forgetting old ones â€” no catastrophic reset.

ðŸŒ–

RSI self-correction

The recursive self-improvement substrate â€” every improvement proposer proven to fire, gated so it can only make the served model better.

ðŸ§

AGI reasoning & agents

Grounded decision reasoning, a what-if oracle, and autonomous experiment-discovery agents â€” analysis, not just prediction.

ðŸ’

LLM + RAG + knowledge graph

A cognitive interface over retrieval + a neuro-symbolic knowledge graph â€” answers grounded in the customer's own data, or NULL.



Physics-informed digital twin

Physics-residual features + a physics rule engine keep predictions consistent with how the equipment actually behaves.



Gates & guards

Promotion, per-task no-regression, transfer, quality, drift, residency & authorization gates â€” the layer that makes trust structural.



Generative studio + KPI attestation

Grounded generative tooling and cryptographically-attested KPIs â€” reporting you can hand to an auditor.



Machine-checked ledger (82/82)

Every capability + gate is re-verified against live code & data as a pre-release gate â€” and **27 checks now RUN the real gate/proposer at runtime**, proving **both halves of self-correction**: the gates block a regression AND every proposer fires on its trigger. Breadth you can audit, proven to work.

Why this compounds for the long term: a static model decays; Ryedore's layers form a loop that **gets better every cycle and can't regress**, runs **on your own infrastructure with your data never leaving**, spans many industries from one foundation, and is **auditable end to end**. The value isn't a single number today â€” it's a platform engineered to widen its lead over time while remaining verifiable and under your control.

Who we measure against, and exactly how.

The field

- **Li 2018 (DCNN):** published academic RUL SOTA “RESS journal.”
- **Chen 2020 (LSTM+attention):** variable-condition RUL “faithful code port.”
- **Mo 2023 (NAS-transformer):** current published var-cond SOTA (full arch-search budget).
- **IsolationForest / One-Class SVM / LOF / Autoencoder:** the standard anomaly detectors a customer would otherwise run.
- **RandomForest / XGBoost:** strong classical ML for tabular fault classification.
- **Chronos-T5:** Amazon's pretrained large general-purpose time-series model (zero-shot forecasting).

Protocol & provenance

- **Public datasets only:** CMAPSS FD001“FD004 (NASA), UCI Hydraulic (CC-BY), ai4i2020 “no customer data.”
- **Held-out test:** official one-per-engine RUL protocol; stratified splits for classification.
- **Honest aggregation:** multi-seed mean $\pm$ std *and* ensemble “no lucky-seed cherry-picking.”
- **Isolated harness:** read-only over results, writes only to its out-dir, degenerate-metric guards.
- **Signed:** RSA-PSS/SHA-256, one canonical key 2bc036d1“; provenance + pinned library versions

Why these four tasks: each maps to a real maintenance decision, not a leaderboard vanity metric. **RUL** drives spare-parts stocking and downtime scheduling “**anomaly detection** catches incipient failures before they cascade “**fault classification** pinpoints root cause so the right work order goes out “**forecasting** sizes consumables, load and capacity. We benchmark the decisions a customer actually makes “and report the honest result on every one.

THE LONG-TERM CASE

What compounding, verifiable, on-your-infrastructure AI is worth over a deployment's life.



Accuracy compounds, not decays

A static model degrades as equipment and operating conditions drift. Ryedore's gated self-improvement means the model you deploy is the **worst it will ever be** — every promoted cycle is provably the last, so avoided-failure value grows year over year instead of eroding.



One foundation, every asset class

New assets and new sites **warm-start** from the shared encoder instead of cold-starting. Coverage scales across 24 industries from one core — onboarding the next line doesn't restart the learning curve.



You own the capability

Runs on-prem / air-gapped, data never leaves site, zero third-party model dependencies, open methodology. No black-box cloud lock-in — the accuracy you build stays yours and stays verifiable.



Audit-ready by construction

Signed KPIs, grounded-or-NULL (it never fabricates a number), and a manual-review floor for safety-critical sectors — reporting you can hand to a regulator, insurer or safety case without a translation layer.



Lower ML-ops burden

The platform benchmarks, proposes, retrains, gates and promotes itself. The engineering labour a closed platform bills as ongoing services is **structural** here â€” the loop does it, gated so it can't ship a regression.



Trust you can re-check anytime

Every headline number is signed and reproducible on public data. Procurement, an independent expert, or your own team can verify each claim â€” no trust-us, no un-runnable case study.

Net: a buyer isn't purchasing today's single accuracy number â€” they're purchasing a capability that **widens its own lead every cycle, can't regress, runs under their control, and stays auditable end to end**. Over a multi-year deployment that is a fundamentally different asset from a static, cloud-locked, unverifiable model.

PROVEN RECORD

The numbers only move one way. Each gain is signed and dated.

Variable-cond RUL Â· FD002 (RMSE â†“)

26.7 **17.85**

arch ladder: scratch â†’ Chen â†’ NAS

Anomaly $\hat{\cdot}$ FD004 (AUC-ROC $\hat{+}$)

0.807 **0.989**

pseudo-labels $\hat{+}$ real-label flywheel

Classification $\hat{\cdot}$ Hydraulic (F1 $\hat{+}$)

0.848 **0.959**

single model $\hat{+}$ family bake-off

Single-cond RUL $\hat{\cdot}$ FD001 (RMSE $\hat{+}$)

14.95 **13.39**

BiLSTM $\hat{+}$ DCNN z-score + ensemble

THE MOAT

A self-improving system that is engineered
so a regression can never ship.

```
flowchart LR
    BM["Signed benchmark  
(current best)"] -->|signature-gated bridge| BL["RSI baseline"]
    BL --> DR["Drift  $\hat{\cdot}$  new labels  $\hat{\cdot}$   
calibration  $\hat{\cdot}$  physics"]
    DR --> PR["Proposer:  
candidate change"]
    PR --> BL
    PR --> TR["Train candidate  
(sandboxed)"]
    TR --> GT["Per-task + baseline  
NO-REGRESSION gate"]
    GT -->|better on EVERY head| PROM["Promote  $\hat{\cdot}$ "]
    GT -->|any head worse| REJ["Reject  $\hat{\cdot}$  keep current  $\hat{\cdot}$ "]
```

```
PROM --> MON["Post-promotion monitor  
(fail-closed)"]  
MON --> BM
```

The never-degrade gate stack “a regression can't ship

- **Five gates on every promotion** “ per-task no-regression • measured-baseline AUC/F1 • calibration (ECE) • OOD-robustness • worst-cohort fairness • serving-budget “ each **wired into the promotion path, with its input computed, and proven to BLOCK a regression at runtime** (not merely present).
- **Proven on a live retrain:** a genuinely-degraded base model that stayed *accurate* (AUC 0.996) but became mis-calibrated (ECE 0.001±0.030) was **rolled back** and the served model byte-restored “ the old accuracy-only check would have shipped it.
- **Atomic staged swap:** a candidate is gated *before* it reaches the served path “ no window where a request could load an un-gated model.

Always improving

- **all self-correction proposers** proven to fire on their triggers.
- **Data flywheel:** real operator failures feed supervised retraining “ the lever that took anomaly FD004 from 0.807 to 0.989.
- **Closed loop:** the platform benchmarks itself, proposes, trains, gates, promotes “ then re-benchmarks. Live now.
- **Machine-checked ledger 82/82** “ every module + gate re-verified; **27 run at runtime** “ gates provably BLOCK a regression and all proposers provably FIRE (behavioral proof, not just presence).

The Online Learning Guarantee a promoted Ryedore model can never be worse than the one it replaces on any head “ so accuracy **compounds upward over time** while regressions are **structurally impossible**

to ship. That is something no static-benchmark competitor can offer, and it is what a customer actually buys.

ROADMAP “ THE HONEST FRONTIER

Where unlimited-compute academic SOTA still leads “ disclosed, and on a measured upward path.

- RUL FD001** 13.39 vs Li 2018 12.6 “ competitive and improving (was 13.80); the last 0.79 RMSE is a plateau needing the paper's exact recipe + extended HPO.
- RUL var-cond** 17.85 / 20.54 beats Li & Chen; FD004 now within 0.54 of Mo 2023 NAS-SOTA (20.0) after the exact published recipe “ the last margin needs a full neural-architecture search (multi-week R&D).
- Forecasting** wins in-domain; beating Chronos zero-shot on unseen domains needs a comparable large general-purpose time-series model “ a disclosed research investment.

The self-correction loop above is precisely how these gaps close over time “ automatically, and without ever risking a regression.

RELIABILITY “ BUILT TO RUN UNATTENDED

Self-correction that improves the model on its own “ engineered so it cannot regress or run away.

RELIABILITY PROPERTY	HOW IT IS PROVEN “ NOT ASSERTED
	Every improvement proposer (drift, calibration, per-task regression, physics, fairness, benchmark-retrain “) has an

RELIABILITY
PROPERTY

HOW IT IS PROVEN " NOT ASSERTED

Every self-correction path actually fires

automated fire-test that trigger them on their real condition " no placebo flags left in the loop.

A regression can never ship

A **named-gate never-degrade stack** guards every promotion " no-regression per task Â· measured-baseline AUC/F1 Â· calibration (ECE) Â· OOD-robustness Â· worst-cohort fairness Â· serving-latency budget Â· data quality " each wired into the promotion path, with its input computed, and runtime-proven to **BLOCK**. Confirmed on a **live base retrain**: a still-accurate but mis-calibrated model was rolled back and the served model byte-restored. Candidates are atomically swapped only after gating (no serving window), and even online updates go through the same gate.

It only improves toward a real target

Self-improvement is gated to a cryptographically-signed benchmark baseline; a tampered or forged number is **refused** " the loop cannot be tricked into "improving" toward a fake metric.

A live promotion " proven, not claimed

A candidate retrain measured AUC-PR 0.8351 â†’ 0.8415 and was promoted *because* it passed every gate; a worse candidate is rejected and the current model kept.

A human stays in control

A single kill-switch halts all self-improvement within **~10 seconds**; it fails *open* on infrastructure issues, so the served model is never wedged or blocked.

The system re-checks itself before release

An **82/82** machine-checked ledger re-verifies every module, gate and guard against live code + data as a pre-release gate " a drifted or hollow claim fails the gate by design.

Why this matters operationally: the platform can be **left running** " it keeps getting better on its own, and the same machinery that lets it improve makes a regression **structurally impossible to deploy**. That is the difference between a model you have to babysit and a capability you can trust in production.

Independently verifiable · safe to share: every headline number is reproducible on public data and RSA-signed; a separate signed audit bundle (checksums, public key, step-by-step guide) is available to auditors on request. This document uses **public benchmark datasets only** " no customer data, no proprietary weights, no internal architecture " so it is safe to share externally.

Ryedore predictive-maintenance platform Â· competitive benchmark audit Â·
public datasets Â· RSA-PSS/SHA-256 signed Â· self-improving Â· regression-proof
Â· wins shown, frontier disclosed Â· 2026-08-16