A Deep Learning-Based Framework for Predicting Remaining Useful Life of Nuclear Safety-Grade DI Cards Using Dynamic Data Synthesis

Ji Hun Park[a], Young Ho Chae[a], and Chang Hwoi Kim[a]
[a] Advanced Instrumentation and Control Division, Korea Atomic Energy Research Institute, Daejeon, Republic of Korea, jihun2992@kaeri.re.kr

Abstract: Digital input (DI) cards in nuclear reactor protection systems show a cliff-edge degradation behavior. Their health is not externally observable, and their protective function is lost abruptly once an internal margin is exhausted. Prognostics and health management (PHM) of such safety-grade equipment is limited by four factors: scarce real failure data, regulatory restrictions on intrusive sensing, the unrealistic fixed-temperature assumption of conventional models, and the absence of uncertainty information in point estimators. This paper presents an end-to-end framework for predicting the remaining useful life (RUL) of the POSAFE-Q NI-D23Q DI card that addresses all four. An eight-month accelerated aging campaign measured 32 channels at 100 °C and 1 Hz and identified the rising threshold voltage as the degradation indicator. Among the channels that failed, CH16 reached the 10.2 V failure criterion at 146 days and was adopted as the physical anchor. Converting this life to a 25 °C operating condition through the Arrhenius law gives a conservative value of 12.49 years. An Arrhenius-Miner cumulative damage model combined with seven-dimensional Latin Hypercube Sampling (LHS) then synthesizes 10,000 virtual units under realistic dynamic temperature trajectories. A Temporal Convolutional Network (TCN) that consumes a 24-hour window and 15 multi-scale features is trained as a five-seed deep ensemble with Conformal Quantile Regression (CQR). It produces an RUL estimate together with a 95 % prediction interval that carries a finite-sample coverage guarantee, separate aleatoric and epistemic uncertainties, and an automatic four-level out-of-distribution (OOD) classifier. On 122,861 held-out test samples the adopted model attains a mean absolute error of 0.167 years (61.0 days), a prediction interval coverage probability of 94.8 % at a mean interval width of 0.70 years, and a PHM08 prognostic score of 1.076. Against a direct physics projection, the model lowers the mean absolute error by 31 % and largely removes the unsafe over-prediction bias. Prediction precision improves steadily as the card approaches end of life. A streaming variant trained with Graceful Augmentation gives stable predictions from the first operating hour and is delivered through a real-time maintenance decision-support interface.

1. INTRODUCTION

Digital instrumentation and control systems are increasingly deployed in nuclear power plants, but their failure behavior differs fundamentally from that of analog equipment. A digital module such as the reactor-protection-system digital input (DI) card shows a cliff-edge characteristic. Degradation produces no externally observable symptom, and the protective function is retained until an internal margin is exhausted, at which point it is lost abruptly. Periodic surveillance confirms health only at the moment of the test and cannot capture the gradual degradation that accumulates between tests.

Prognostics and health management (PHM) offers a way forward [1]. By analyzing condition information, PHM detects anomalies early and predicts the remaining useful life (RUL), which enables a shift from time-based to condition-based maintenance. For safety-grade equipment this shift promises simultaneous gains in safety, economy, reliability, and the reduction of human error. Applying PHM to nuclear safety-grade DI cards, however, faces four practical constraints.

- **Data scarcity.** The high reliability of safety-grade equipment makes real run-to-failure data extremely rare.
- **Regulatory restriction.** Intrusive in-circuit measurement and the addition of new sensors are strictly prohibited, so only externally available signals may be used.

• **Unrepresented dynamics.** A fixed operating-temperature assumption generalizes poorly to the thermal variability of real cabinets.
• **Absence of confidence.** A point estimate provides no uncertainty quantification and is ill-suited to conservative safety decisions.

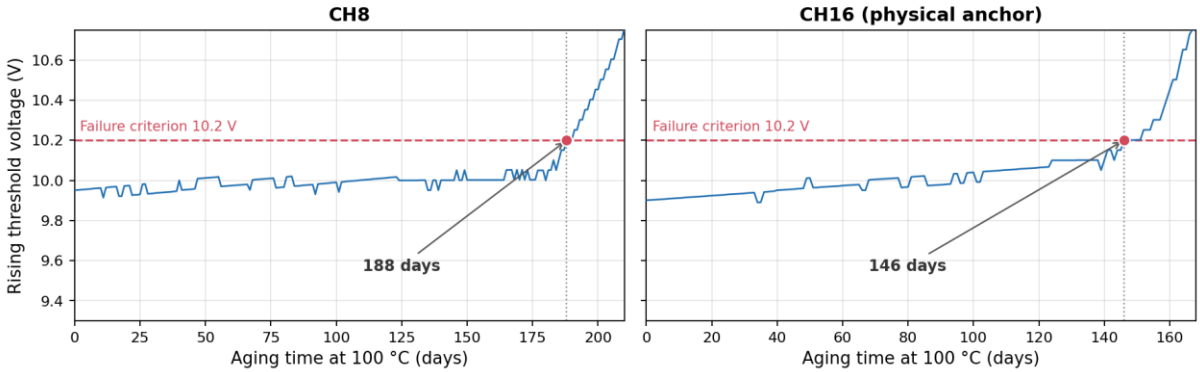
This work resolves the four constraints with one integrated framework and makes four corresponding contributions. First, it builds a physics-anchored synthetic dataset from an accelerated aging campaign and the Arrhenius law, which overcomes the scarcity of failure data. Second, the predictor uses only the cabinet-external temperature and the operating time, so no hardware change is required. Third, the data synthesis reproduces realistic dynamics, including daily and annual cycles, short-term and long-term drift, and occasional thermal excursions caused by heating, ventilation, and air-conditioning (HVAC) faults. Fourth, the output is uncertainty-aware: it combines a point estimate, a 95 % prediction interval guaranteed by Conformal Quantile Regression, and automatic OOD classification. The remainder of the paper is organized as follows. Section 2 develops the physics anchor, Section 3 the data synthesis, and Section 4 the prediction model. Section 5 reports the test-set results, Section 6 describes the decision-support interface, and Section 7 concludes.

2. ACCELERATED AGING TEST AND PHYSICS ANCHOR

2.1. Degradation Mechanism and Indicator

The POSAFE-Q NI-D23Q is a Class-1E qualified [2] 32-channel DC digital input module. Each channel discriminates a field signal as it rises from 0 V to 24 V. The voltage at which the logic state transitions, called the rising threshold voltage, quantifies the health of the internal discrimination circuitry. The dominant aging mechanism is thermal degradation of the input optocoupler, whose current transfer ratio falls under accumulated thermal stress and drives the threshold voltage upward; this mechanism sets the activation energy used in Section 2.3. Because safety-grade equipment operates for decades, degradation data cannot be obtained under normal conditions. Degradation was therefore accelerated in a constant-temperature chamber held at 100 °C, and the voltage signals of all 32 channels (designated CH0 to CH31) were logged continuously at 1 Hz from August 2024 to April 2025, a period of about eight months. The rising threshold voltage drifts monotonically with time and has a high signal-to-noise ratio, so it was selected as the principal degradation indicator. Figure 1 presents two representative channels. Each channel holds a nearly flat voltage for most of the test and then rises sharply through the 10.2 V criterion at the knee point, which confirms the expected wear-out behavior.

Figure 1: Measured threshold-voltage degradation of CH8 and CH16.
Monotonic threshold-voltage drift to the 10.2 V knee point

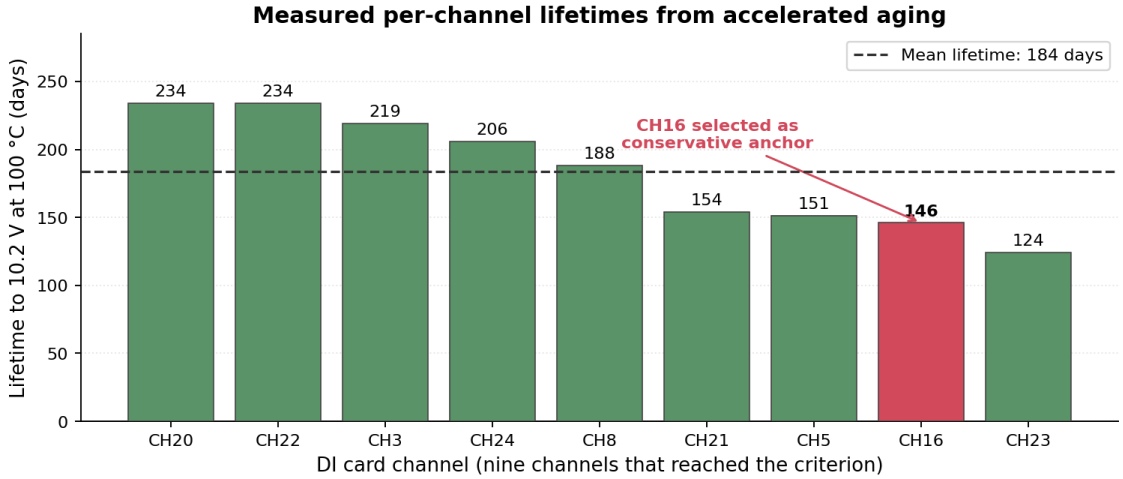


2.2. Failure Criterion and Conservative Life Anchor

The instant at which the rising threshold exceeds 10.2 V defines the functional end of life. This instant is the knee point, where the slow drift turns into a rapid exponential increase. Of the 32 aged channels,

nine reached the 10.2 V criterion within the campaign window. Their lifetimes ranged from 124 to 234 days, with a mean of 184 days (Figure 2). Among them, CH16 reached the criterion at 146 days, near the conservative lower end of this range, and was adopted as the reference anchor. Converting its 100 °C life to a nominal 25 °C operating condition through the Arrhenius acceleration factor gives a conservative life of 12.49 years, which is used as the ground-truth end of life for model training.

Figure 2: Measured lifetimes of the nine failed channels.



2.3. Arrhenius-Miner Damage Model

Thermally activated degradation is described by the Arrhenius acceleration factor relative to the 100 °C reference. This factor converts any operating temperature into an instantaneous damage rate:

$$AF(T) = \exp[\ (E_a / k_B) \cdot (\ 1/T_{ref} - 1/T\ )\ ] \qquad (1)$$

Here the activation energy is E_a = 0.44 eV, a literature value for optocouplers, the Boltzmann constant is k_B = 8.617×10^{-5} eV/K, and the reference temperature is T_ref = 373.15 K. The equivalent age accumulated along a temperature trajectory T(τ), together with the Miner linear cumulative damage [6], is

$$t_{eq}(t) = \int_0^t AF(T(\tau))\, d\tau\ , \quad D(t) = t_{eq}(t) / \tau_{fail} \qquad (2)$$

where τ_fail is 146 equivalent days at 100 °C, the value fixed by the CH16 anchor, and failure occurs when D(t) reaches 1. The threshold voltage is modelled as an exponential function of equivalent age, v = a·exp(b·t_eq) + c, with coefficients (a, b, c) equal to (0.269, 9.45×10^{-3}, 9.131). These coefficients fix the voltage at 9.4 V when t_eq is 0 and at 10.2 V when t_eq is 146, which matches the measured CH16 curvature.

3. PHYSICS-BASED DYNAMIC DATA SYNTHESIS

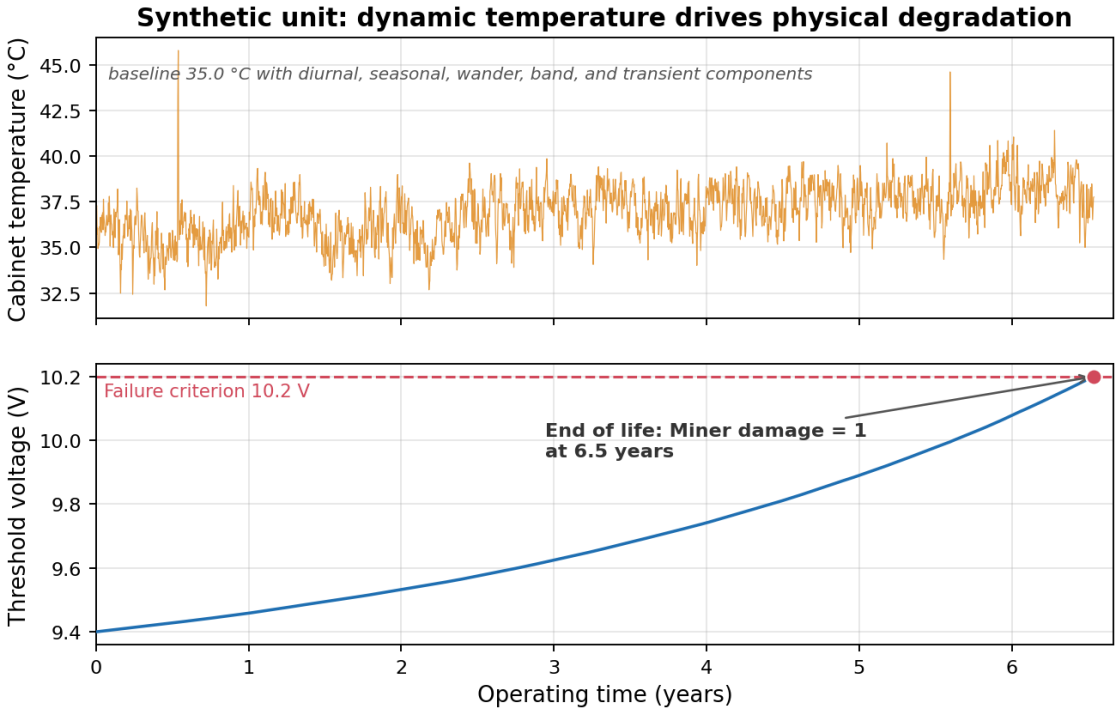
A fixed 100 °C test is efficient for analyzing the degradation mechanism, but it cannot represent the dynamic environment of a real plant. To generate enough realistic training data, 10,000 virtual units are synthesized at hourly resolution. The synthesis combines a dynamic temperature model with the Arrhenius-Miner physics of Section 2.

3.1. Dynamic Temperature Model

Each unit follows a cabinet temperature built from eight physically motivated components. These are a baseline with a slow linear trend, a diurnal sinusoid, a seasonal sinusoid, a low-frequency Ornstein-Uhlenbeck wander (σ = 0.6 °C, τ = 45 days), a first-order autoregressive short-term fluctuation, short

and long thermal excursions that represent HVAC faults, a first-order cabinet thermal lag, and sensor noise. The sum is clipped to the qualified envelope of 18 to 60 °C. Figure 3 shows one representative unit. The dynamic temperature drives the threshold voltage to 10.2 V exactly when the Miner damage reaches 1.

Figure 3: Temperature-driven degradation of a representative synthetic unit.



3.2. Seven-Dimensional Latin Hypercube Sampling

Unit-to-unit variability is introduced by sampling seven temperature parameters with Latin Hypercube Sampling (LHS) [10] under random-CD optimization. LHS guarantees uniform stratified coverage of the parameter space and does so far better than independent random sampling. No intrinsic lifetime scatter term is added, so lifetime variation arises entirely from the sampled temperature history. This keeps the mapping from environment to life fully deterministic and physically interpretable. Table 1 lists the seven parameters and their ranges, which were chosen to span the conditions seen across plant locations, seasons, and ventilation states. The first two parameters set the temperature level, the next two set its dynamics, and the last three set the periodic and random fluctuations. The derived baseline temperature, equal to the room temperature plus the self-heating rise, spans 24 to 47 °C with a mean of 35.5 °C.

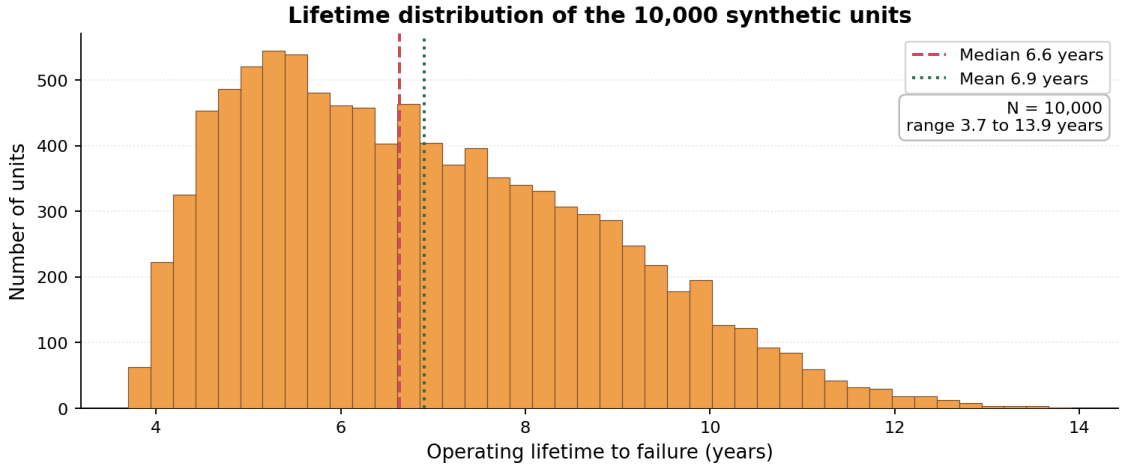
Table 1: The seven LHS-sampled temperature parameters (10,000 units, seed 42).

Parameter	Range	Unit	Physical role
Room temperature	21 to 27	°C	Ambient temperature
Self-heating rise	3 to 20	°C	Internal heat rise
Trend	−0.2 to 0.7	°C/yr	Long-term drift
Thermal time constant	6 to 12	h	Cabinet thermal lag
Short-term noise	0.8 to 1.6	°C	HVAC fluctuation
Diurnal amplitude	0.3 to 1.5	°C	Daily cycle
Seasonal amplitude	0.2 to 0.7	°C	Annual cycle

Each trajectory is integrated until the Miner damage reaches 1, which yields the population lifetime distribution in Figure 4. The median life is 6.6 years, the mean is 6.9 years, and the range is 3.7 to 13.9 years. Most units run hotter than the 25 °C nominal condition, so their real-time lifetimes are shorter than the 12.49-year anchor while remaining physically consistent with it, because every unit

accumulates exactly 146 equivalent days at failure. The 10,000 units are split at the unit level into 7,000 for training, 1,500 for validation, and 1,500 for testing, so that no unit appears in more than one split.

Figure 4: Lifetime distribution of the 10,000 synthetic units.



4. RUL PREDICTION MODEL

4.1. Architecture Selection

DI-card degradation is encoded in the accumulated thermal-stress history rather than in any instantaneous signal, so capturing temporal context is essential. Given the complete future temperature, the remaining life would follow deterministically from the Arrhenius-Miner model. The task is difficult precisely because only the past, cabinet-external temperature is observable, so the network must infer the unobserved future thermal load from causal context. Seven representative sequence architectures were optimized under one identical protocol, using point regression at daily resolution. The seven were a deep neural network (DNN), a long short-term memory network (LSTM), a bidirectional LSTM (Bi-LSTM), a convolutional gated recurrent unit (CNN-GRU), a Temporal Convolutional Network (TCN) [3], a Transformer, and N-BEATS. They were compared by mean absolute error (MAE), root-mean-square error (RMSE), the asymmetric PHM08 prognostic score [7], and the fraction of predictions within 5 % and 10 % of the true value. As Table 2 shows, the TCN gave the best accuracy on every error metric and was therefore promoted to hourly resolution.

Table 2: Daily architecture benchmark on the test set, using point regression under an identical protocol.

Architecture	MAE [yr]	RMSE [day]	PHM08	Within 5 %	Within 10 %
DNN	0.237	121.4	1.583	45.5	73.2
LSTM	0.233	120.1	1.563	46.8	74.4
Bi-LSTM	0.230	119.0	1.517	47.4	75.1
CNN-GRU	0.230	118.6	1.560	47.8	75.5
TCN (selected)	0.228	118.1	1.532	48.5	76.0
Transformer	0.240	120.9	1.595	42.9	71.7
N-BEATS	0.230	118.8	1.559	47.8	75.4

The TCN has three advantages for this task. It processes the whole input window in parallel, unlike the sequential recurrence of an LSTM. It learns patterns at several time scales, namely one day, one week, one month, and three months, within a single model. It also supports ensembling naturally, which the framework uses for uncertainty quantification.

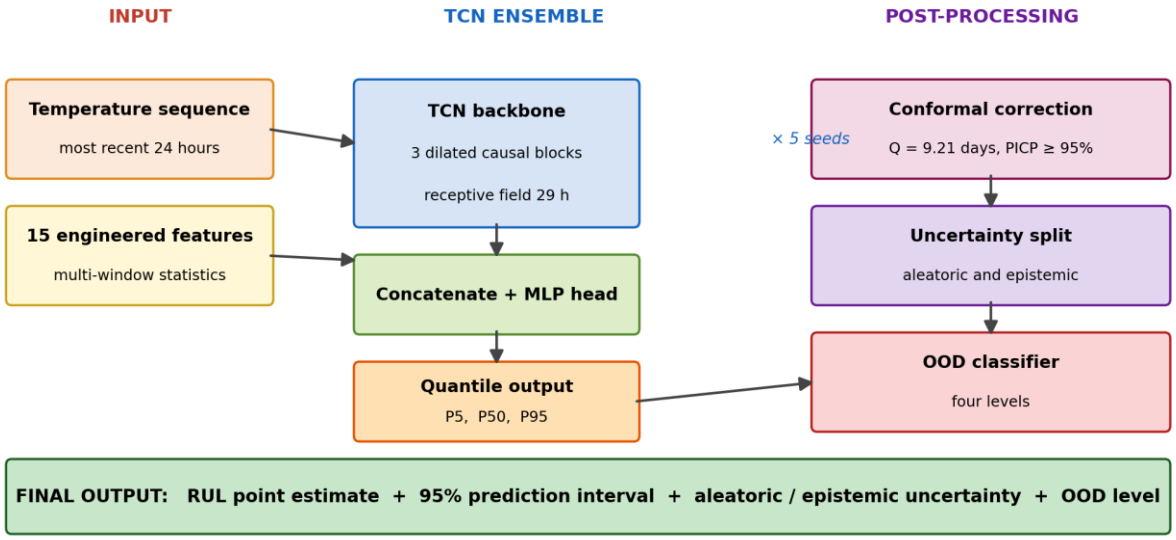
4.2. TCN Ensemble and Input Features

The adopted model consumes a 39-dimensional input. This input consists of a 24-element raw sequence, the most recent 24 hours of temperature, and 15 engineered features. The features aggregate temperature statistics (mean, minimum, and maximum) over four windows of 24, 168, 720, and 2160 hours, and add three cumulative history variables, namely the operating time, the equivalent time, and the lifetime-mean temperature. Table 3 lists them. The TCN backbone uses dilated causal convolutions whose 29-hour receptive field covers the 24-hour window. Its pooled output is concatenated with the 15 features and passed to a multilayer perceptron head that produces three quantile outputs, namely P5, P50, and P95. Figure 5 shows the full pipeline. The two inputs enter the TCN backbone and the feature concatenation on the left, five independent seeds form the ensemble in the center, and the conformal correction, the uncertainty split, and the OOD classifier on the right turn the raw quantiles into the final calibrated output.

Table 3: The 15 engineered features.

Group	Features	Definition
Multi-window (12)	Temperature mean, min, max	over 24, 168, 720, and 2160 hours
Cumulative (3)	Operating time	total elapsed hours
	Equivalent time	Arrhenius-weighted hours
	Lifetime-mean temperature	running mean temperature

Figure 5: TCN ensemble architecture and post-processing pipeline.



4.3. Conformal Quantile Regression

Instead of a single point, the model outputs a predictive range. Each TCN is trained with the pinball quantile loss at q values of 0.05, 0.50, and 0.95, which directly produces a lower, a median, and an upper estimate:

$$L_q(y, \hat{y}_q) = max[\, q(y - \hat{y}_q), (q - 1)(y - \hat{y}_q) \,] \quad (3)$$

Pinball training alone gives no finite-sample coverage guarantee. Split Conformal Quantile Regression [4], [8] is therefore applied. On the validation split the non-conformity score $E_i = \max(\hat{y}_{0.05} - y_i, y_i - \hat{y}_{0.95})$ is computed, and a single correction Q is set to its empirical $(1 - \alpha)$ quantile. Every interval is then widened to $[\hat{y}_{0.05} - Q, \hat{y}_{0.95} + Q]$. This widening guarantees a prediction interval coverage probability (PICP) of at least $1 - \alpha$ under exchangeability. For the adopted model Q is 9.21 days, so on average 95 of every 100 true RUL values fall inside the corrected interval.

4.4. Uncertainty Decomposition and OOD Detection

Five TCNs are trained independently from different seeds, which forms a deep ensemble [5]. Their disagreement measures how uncertain the model is. Following the standard decomposition [9], two uncertainty types are separated:

$$\sigma_{alea} = \text{mean}_{seeds}[(P95 - P5)/2] \,, \quad \sigma_{epi} = \text{std}_{seeds}[P50] \tag{4}$$

Because the generator adds no label noise, σ_alea here does not represent measurement scatter. Instead it captures the irreducible ambiguity that remains when only the past cabinet-external temperature has been observed and the future thermal load is unknown. The epistemic uncertainty σ_epi is the reducible model uncertainty that shrinks as data accumulate. The input is graded automatically into four OOD levels by comparing σ_epi against percentiles of the validation distribution (Table 4). This grading flags when the model is extrapolating beyond its training distribution, so overconfident predictions are avoided in safety decisions.

Table 4: The four-level OOD classifier, based on validation percentiles of the epistemic uncertainty.

Level	σ_epi (days)	Meaning	Recommended action
normal	at most 11.6	in distribution	use the prediction
borderline	11.6 to 14.4	near the boundary	monitor
ood_yellow	14.4 to 22.0	low reliability	verify and schedule
ood_red	above 22.0	clearly out of distribution	reject the prediction

4.5. Streaming Inference and Graceful Augmentation

In deployment the model must predict from the first operating hour, when the 24-hour buffer is only partially filled. The missing positions are filled with the observed mean, a method called B_mean padding. A model trained only on full 24-hour windows has never seen such partially padded inputs, however, and would raise false ood_red alarms during the first day. Graceful Augmentation closes this gap between the training and inference distributions. During training, a random effective length is drawn for each batch, and the leading positions are overwritten with the same B_mean padding that inference uses. The adopted model, denoted M3, uses 15 features. The standard-deviation and slope statistics of the full-window variant are removed because they are ill-defined under heavy padding. As a result, the model is robust to partial inputs and gives stable predictions from the first operating hour without any warmup phase.

5. RESULTS AND DISCUSSION

All models are evaluated on the 1,500 held-out units, which yield 122,861 test samples. The conformal correction Q is fixed on the validation split before testing.

5.1. Comparison with Baselines and Variants

Moving from daily to hourly resolution and adding the five-seed CQR ensemble reduced the test MAE from 0.228 years, for the best daily point model, to 0.156 years. Table 5 compares the hourly ensemble variants. The 23-feature variant with a 24-hour window is the most accurate, but it raises false ood_red alarms on partially padded streaming inputs. The adopted M3 model uses 15 features and trades a small amount of accuracy, 0.167 against 0.156 years, for correct streaming OOD behavior through Graceful Augmentation. It is therefore selected for deployment.

Table 5: Hourly TCN ensemble variants on the test set. All variants use a 24-hour window and five seeds.

Model	Features	MAE [yr]	RMSE [day]	PHM08	PICP [%]	Interval width [yr]
24-hour ensemble	23	0.156	92.8	0.985	94.8	0.67

24-hour, single window	8	0.174	99.9	1.126	94.6	0.72
M3 ensemble (adopted)	15	0.167	97.4	1.076	94.8	0.70

To confirm that the learned model adds value beyond the underlying physics, it is compared against direct projections built from the same Arrhenius-Miner equations that generated the data. Both the projections and the model use the same past observations, the accumulated equivalent time and the recent temperature, and neither can observe the future temperature. The unknown future is therefore a common limit and does not by itself favor either approach; what differs is how the future is estimated. Each physics baseline assumes a fixed future, taking the current temperature, the average past damage rate, or the lifetime-mean temperature to persist until failure. As Table 6 shows, the best physics baseline reaches a mean absolute error of 89.1 days against 61.0 days for the M3 model, a reduction of 31 %, and every physics projection over-predicts the remaining life by 55 to 75 days, the unsafe direction of error, against only +3.4 days for the model. This over-prediction has two physical causes. The sampled units warm slightly over time, so the true future damage rate exceeds the past average that the projection assumes. In addition, because the damage rate rises exponentially with temperature, unforeseen high-temperature episodes inflict disproportionate damage that a smooth projection cannot anticipate. The model learns both effects from the training population and so estimates the future damage trajectory better than any fixed-future assumption.

Table 6: The M3 model against direct physics projections (test set, days).

Model	MAE	RMSE	PHM08	Within 10 %	Mean error
Physics, current temperature	106.1	173.5	65.9	74.3	+54.7
Physics, average rate	89.1	139.6	4.53	77.2	+69.3
Physics, lifetime-mean temp.	91.9	143.7	4.98	75.3	+74.8
M3 ensemble (this work)	61.0	97.4	1.08	92.5	+3.4

5.2. Prediction Accuracy

Figure 6 shows predicted against actual RUL for the adopted model. The identity diagonal is recovered across the full range of 0 to 5,000 days, with no systematic drift. The mean absolute error is 61.0 days and the root-mean-square error is 97.4 days. The residual distribution is approximately symmetric and unimodal, and its mean error is only +3.4 days, which indicates a slight and deliberately conservative tendency. Figure 7 overlays the prediction and the 95 % conformal interval on the ground truth for nine randomly sampled test units. In every unit the interval contains the true RUL throughout the operating history, and it narrows steadily as the unit approaches end of life.

Figure 6: Predicted against actual remaining useful life.

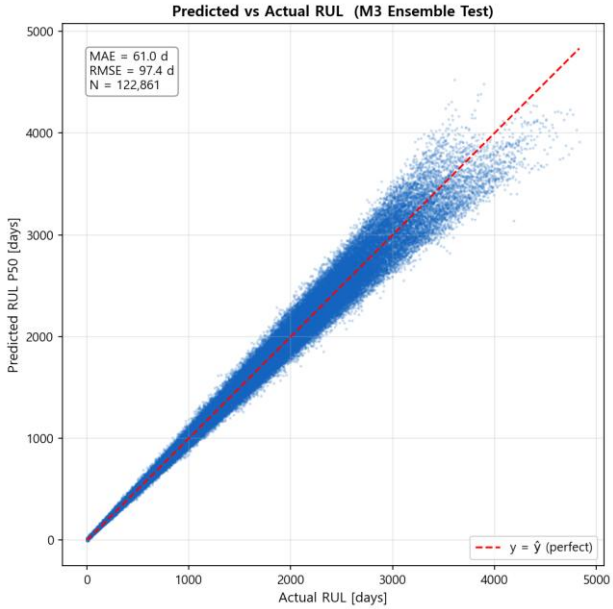
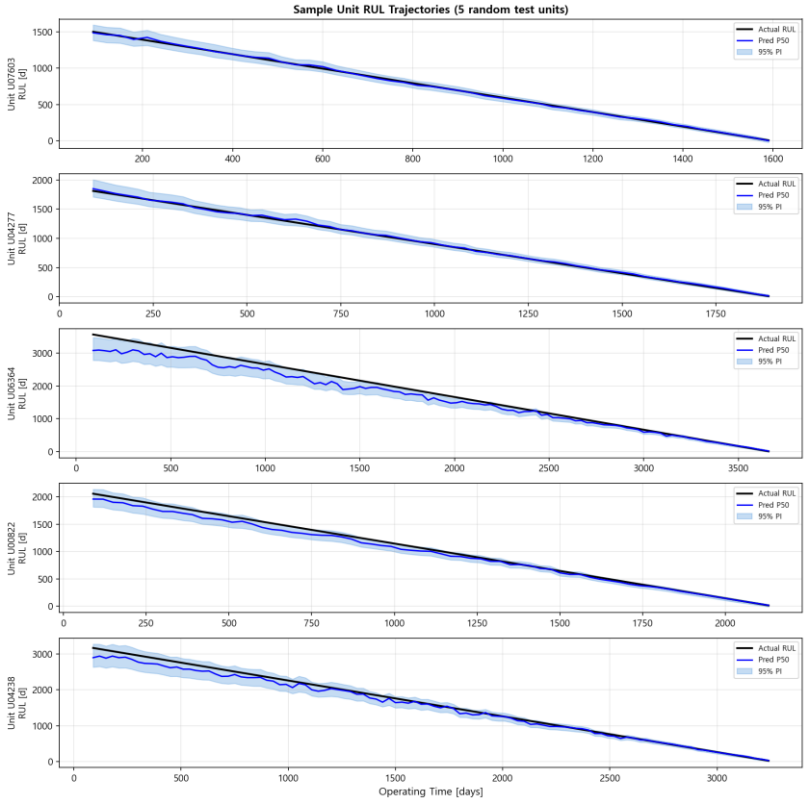


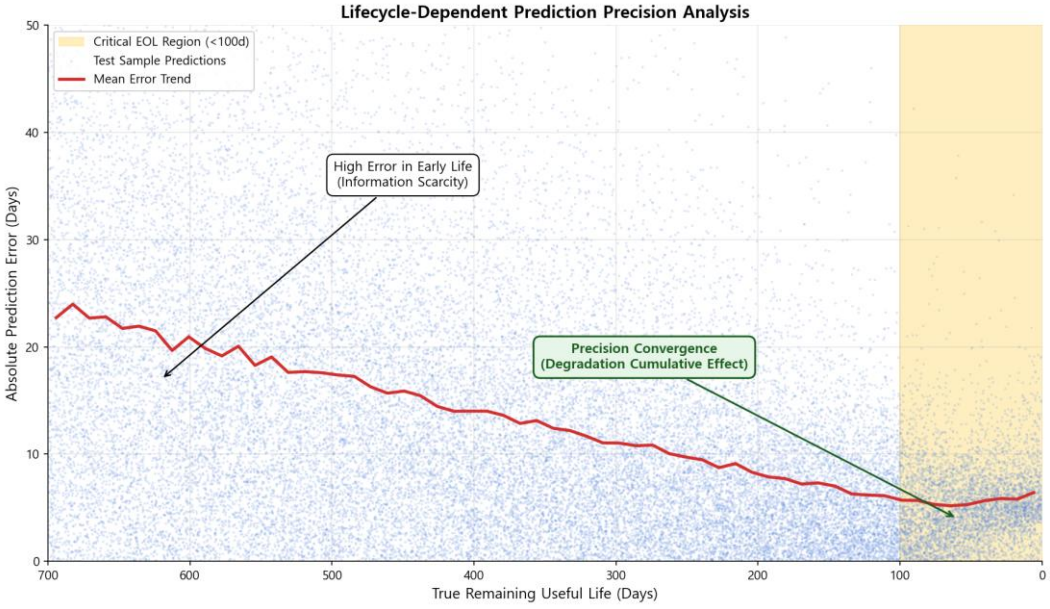
Figure 7: Prediction intervals for nine sample test units.



5.3. Lifecycle-Dependent Precision

Accuracy near end of life matters most in practice. Figure 8 plots the absolute error against the true RUL, with the axis reversed so that the critical region below 100 days is on the right. Early in life, limited window history and the spread of the population lifetime distribution produce large errors. The error then decreases steadily as degradation accumulates, and it converges to high precision near end of life. The mean absolute error falls from 74 days when the RUL exceeds 500 days to 5.6 days within the critical region below 100 days. Precision thus concentrates exactly where maintenance decisions are made.

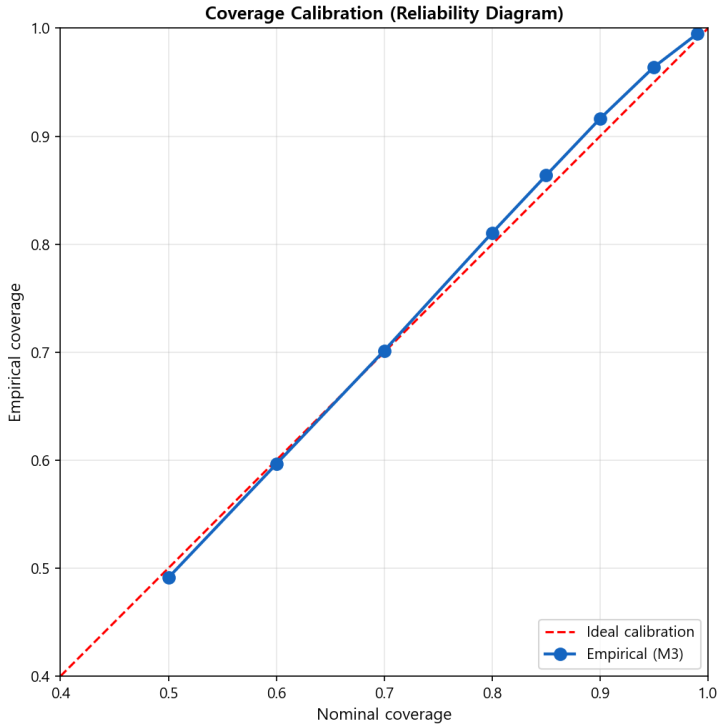
Figure 8: Lifecycle-dependent prediction precision.



5.4. Uncertainty Calibration

Figure 9 is the reliability diagram, which compares the empirical coverage with the nominal coverage. The curve tracks the diagonal across the full range. At the nominal 95 % level the empirical PICP is 94.8 %, which lies within the finite-sample slack of the conformal guarantee. The model therefore reports trustworthy intervals rather than overconfident point estimates.

Figure 9: Conformal interval reliability diagram.



5.5. Performance Summary

The adopted model is accurate and well calibrated at the same time. Its mean absolute error of 0.167 years is about 61 days over a multi-year life, and 92.5 % of all predictions fall within 10 % of the true RUL. The PHM08 score of 1.076 reflects the small conservative bias, which is the preferred direction of error for a safety decision because it favors early rather than late maintenance. The empirical coverage of 94.8 % matches the 95 % target set by the conformal correction of 9.21 days, and the three epistemic-uncertainty thresholds of 11.6, 14.4, and 22.0 days define the four out-of-distribution levels applied during operation.

5.6. Limitations

Several caveats qualify these results. First, training, calibration, and testing all draw on synthetic data from a single generator, so the metrics measure how faithfully the network inverts that generator rather than its accuracy in the field. The accelerated aging campaign provides only constant-temperature data, which lies outside the dynamic training envelope and cannot exercise the dynamic-input pathway of the model, so validation on field temperature histories with confirmed failures remains the essential next step. Second, because the generator adds no intrinsic scatter, the conformal coverage guarantee holds by construction under the exchangeability of the calibration and test data; field deployment introduces distribution shift, under which the guarantee can degrade and online recalibration becomes necessary. Third, the out-of-distribution thresholds are derived from the synthetic validation distribution, so they detect departures from the synthetic dynamics rather than from true field conditions, and any input that lacks realistic temperature dynamics is flagged accordingly. Fourth, the 12.49 year reference rests on a single channel, a single 10.2 V criterion, and a single assumed mechanism; since the measured channel

lifetimes span 124 to 234 days, the anchor choice materially affects the absolute life scale, and a sensitivity study with multi-channel anchoring is warranted.

6. MAINTENANCE DECISION-SUPPORT INTERFACE

The adopted model is integrated into a real-time decision-support prototype. At each hour a new temperature sample produces a new RUL estimate with its 95 % interval, and prediction starts at the first operating hour without any warmup. The interface draws a forward uncertainty cone that projects the interval to the point where the RUL reaches zero, a four-level OOD color badge, and separate bars for the aleatoric and epistemic uncertainty. A maintenance layer maps the RUL to automatic risk grades, namely normal, replacement recommended, and urgent replacement, with user-adjustable thresholds. All inference runs in PyTorch on a single processor thread and updates each simulated hour, so the prototype can replay recorded or simulated temperature histories for operator training and for what-if analysis.

7. CONCLUSION

This paper presented an end-to-end framework that predicts the remaining useful life of nuclear safety-grade POSAFE-Q NI-D23Q digital input cards together with a calibrated uncertainty. By anchoring an Arrhenius-Miner model on an eight-month accelerated aging campaign and synthesizing 10,000 dynamic virtual units with seven-dimensional LHS, the framework overcomes the scarcity of real failure data while using only the cabinet-external temperature and the operating time, with no hardware change. A 24-hour TCN deep ensemble with Conformal Quantile Regression delivers a point estimate, a 95 % prediction interval, decomposed aleatoric and epistemic uncertainty, and an automatic four-level OOD warning, and Graceful Augmentation enables stable prediction from the first operating hour. On the synthetic test set the model attains a mean absolute error of 0.167 years and a coverage of 94.8 %, with precision improving toward end of life. As set out in Section 5.6, the principal remaining task is validation on field data; future work will also add online conformal recalibration and extend the framework to other instrumentation and control modules.

Acknowledgements

This work was supported by the National Research Foundation of Korea grant funded by the Korea government (Ministry of Science and ICT) under the Nuclear Safety Enhancement Core Technology Development Program for Operating Nuclear Power Plants (No. RS-2022-00144521).

References

[1] M. Pecht, *Prognostics and Health Management of Electronics*, Wiley-Interscience, Hoboken, NJ, (2008).
[2] IEEE Std 323-2003, *IEEE Standard for Qualifying Class 1E Equipment for Nuclear Power Generating Stations*, IEEE, New York, (2003).
[3] S. Bai, J. Z. Kolter and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling", *arXiv:1803.01271*, (2018).
[4] Y. Romano, E. Patterson and E. Candes, "Conformalized Quantile Regression", *Advances in Neural Information Processing Systems*, vol. 32, pp. 3543 to 3553, (2019).
[5] B. Lakshminarayanan, A. Pritzel and C. Blundell, "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles", *Advances in Neural Information Processing Systems*, vol. 30, pp. 6402 to 6413, (2017).
[6] M. A. Miner, "Cumulative Damage in Fatigue", *Journal of Applied Mechanics*, vol. 12, pp. A159 to A164, (1945).

[7] A. Saxena, K. Goebel, D. Simon and N. Eklund, “Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation”, *Proc. International Conference on Prognostics and Health Management*, pp. 1 to 9, (2008).
[8] V. Vovk, A. Gammerman and G. Shafer, *Algorithmic Learning in a Random World*, Springer, New York, (2005).
[9] A. Kendall and Y. Gal, “What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?”, *Advances in Neural Information Processing Systems*, vol. 30, pp. 5574 to 5584, (2017).
[10] M. Stein, “Large Sample Properties of Simulations Using Latin Hypercube Sampling”, *Technometrics*, vol. 29, pp. 143 to 151, (1987).