An Integrated Health Index, Remaining Life Assessment, and Probability of Failure Framework for Critical Electrical Assets

Anish Kumar Pal
Department of Electrical Engineering
Indian Institute of Technology Bombay
Mumbai, India
25m1087@iitb.ac.in

***Abstract*—Unplanned outages of critical electrical assets such as transformers, gas-insulated switchgear (GIS), motors, and power cables impose severe operational, financial, and safety penalties on modern power networks. Conventional condition-assessment practice relies on discrete, periodic inspections that fail to capture the dynamic operational and environmental stresses that accelerate ageing. This paper presents an integrated, microservice-based framework that unifies three complementary analytics: (i) a *Dynamic Health Index (HI) Service* that fuses heterogeneous diagnostic measurements into a single weighted condition score and adjusts it for real-world stress through a Conditional Factor (CF) and a Weibull survival function; (ii) a *Remaining Life Assessment (RLA) Service* that fits a hybrid Weibull–polynomial degradation model to smoothed historical HI trajectories and projects Remaining Useful Life (RUL) by threshold crossing; and (iii) a *Probability of Failure (PoF) Service* that couples survival analysis with Monte Carlo uncertainty propagation to express failure risk over finite horizons. We further describe a partial-discharge (PD) monitoring subsystem that supplies a key diagnostic input through five-minute phase-resolved partial discharge (PRPD) snapshots. The HI engine is realised as a configuration-driven pipeline: JSON-defined parameter rules drive data simulation and scoring, a gradient-boosted regressor (LightGBM) learns the mapping from raw measurements to HI, and the model is exported to the ONNX format for portable, low-latency inference behind a REST API. The resulting platform enables risk-based maintenance planning, prioritisation of ageing assets, and improved capital-expenditure forecasting.**

***Index Terms*—health index, remaining useful life, probability of failure, predictive maintenance, Weibull distribution, partial discharge, condition monitoring, machine learning, gradient boosting, asset management.**

I. Introduction

As electrical infrastructure continues to expand in scale and complexity, ensuring the reliable operation of critical assets—transformers, circuit breakers, motors, and cables—has become increasingly vital. These assets form the backbone of power generation, transmission, and distribution, and their failure can result in substantial operational, financial, and safety consequences. Despite routine maintenance, unplanned downtime still occurs, often due to progressive degradation that remains undetected until it is too late.

To proactively manage reliability, many utilities have adopted *Health Index (HI)* frameworks: data-driven methods that consolidate diverse condition monitoring parameters into a single score reflecting the current health of an asset [1]–[3]. Conventional HI models rely on discrete inspection data such as thermal imaging, vibration analysis, oil testing, partial-discharge measurements, or insulation-resistance readings. While valuable, these assessments are performed at scheduled intervals and do not account for the dynamic operational and environmental stressors that accelerate ageing. Consequently, static HI assessments fall short of delivering accurate forecasts of remaining useful life.

This paper addresses these limitations with a more robust and adaptive approach that integrates both historical condition data and real-time operational context. We describe the design and implementation of a *Dynamic Health Index Service*, and extend it with two downstream analytics that consume HI outputs: a *Remaining Life Assessment (RLA)* service and a *Probability of Failure (PoF)* service. Combining long-term health trends with condition-based adjustment factors yields a more accurate and responsive estimate of asset health and lifespan, enabling:

- enhanced risk-based maintenance planning;
- prioritisation of ageing or high-risk assets;
- better capital-expenditure forecasting; and
- improved system reliability and safety.

The contributions of this paper are: (1) a configuration-driven, machine-learning HI service that decouples domain knowledge (expert weightings and condition rules) from code; (2) a hybrid Weibull–polynomial RLA model that operates on smoothed HI trajectories augmented by a Conditional Factor; (3) a survival-analysis based PoF formulation with Monte Carlo uncertainty propagation; and (4) a partial-discharge monitoring subsystem providing PRPD-based diagnostic inputs. For reproducibility, sensitive and product-specific details from the underlying production system have been deliberately omitted or generalised.

II. Related Work

The HI concept was popularised for power transformers, where weighted scoring of dissolved gas analysis, oil quality, furan content, and electrical tests is used to derive a single comparative condition score [1], [2]. Subsequent work extended HI methodologies to transmission lines, underground cables, and other components using condition-based methods [3], [4], and incorporated operating history, load rate, and ageing into the score [5], [6]. Recent studies apply machine learning to learn the HI mapping directly from condition data, improving scalability and reducing manual recalculation [4], [7].

Reliability and prognostics literature provides the statistical backbone for life estimation. The Weibull distribution [8] is widely used to model time-to-failure, and has been combined with data-driven models for remaining useful life (RUL) prediction in rotating machinery [9], [10]. For failure-risk quantification, survival analysis methods such as the Cox proportional hazards model [11] and random survival forests [12] estimate hazard and survival functions from censored event data. Partial-discharge monitoring is a recognised indicator of insulation degradation in high-voltage equipment [13], with measurement and field-testing practice codified in IEC 60270 [14] and IEEE 400.3 [15]. Our work brings these threads together into a single deployable pipeline built around gradient-boosted regression [16] and portable ONNX inference [17].

III. System Architecture

The platform follows a layered, microservice-oriented architecture so that the three analytics can scale and evolve independently while integrating with an enterprise-level monitoring solution through REST APIs.

1) **Data Acquisition Layer:** collects real-time sensor data (including PD metrics) and historical condition records.
2) **Processing Layer:** cleans, smooths, and structures raw data, and computes derived condition features.
3) **Prediction Layer:** executes the HI, RLA, and PoF models.
4) **Visualisation Layer:** renders results on interactive dashboards (HI trends, projected decay, risk heatmaps).
5) **Notification Module:** raises automated alerts and work orders when degradation or risk thresholds are crossed.

This separation of concerns yields several practical benefits: services can be scaled horizontally according to load, new asset classes or model versions can be onboarded without modifying upstream layers, and failures are isolated to a single service rather than cascading through the platform. Communication over lightweight REST APIs also simplifies integration with existing enterprise monitoring tools. Figure 1 summarises the end-to-end flow, in which the HI service acts as the central producer of time-series condition scores that are subsequently consumed by the RLA and PoF services.

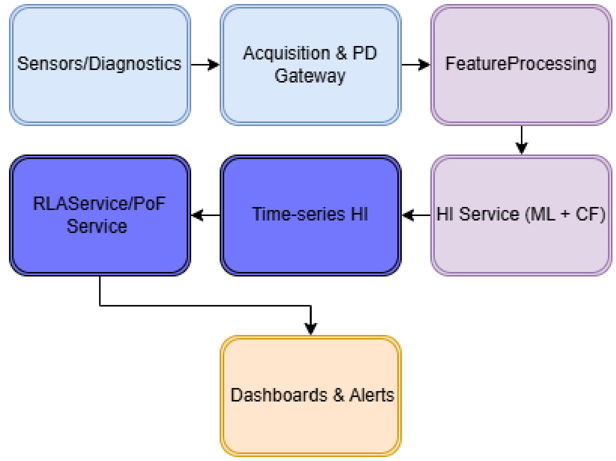

Fig. 1. End-to-end analytics pipeline. The Health Index service is the central producer of condition scores consumed by the RLA and PoF services.

IV. Dynamic Health Index Service

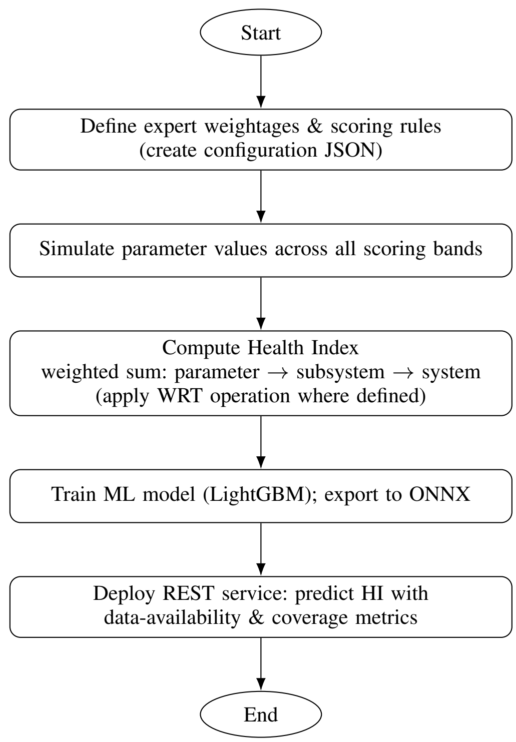

Fig. 2. Processing pipeline of the ML-based Health Index calculation and prediction service.

A. Health Index Formulation

The Health Index is a weighted score derived from multiple diagnostic tests and inspections, each assigned a numeric score and an importance factor. Typical inputs include partial-discharge measurements, infrared thermography, sheath current monitoring, visual inspections, and grounding-resistance testing. Each test result i is mapped to a score S_i and weighted by its significance W_i for the relevant asset class (e.g. cable, joint, termination, manhole, or duct bank).

The *component* health index $\%\mathrm{HI}_C$ for a major asset part is

$$\%\mathrm{HI}_C = \frac{\sum_i W_i \left(\frac{S_i}{S_{\max,i}}\right)}{\sum_i W_i} \times 100, \tag{1}$$

where $S_{\max,i}$ is the maximum attainable score for test i. Once each component index is known, the overall *system* health index $\%\mathrm{HI}_S$ aggregates the worst-case component values across groups:

$$\%\mathrm{HI}_S = \frac{\sum_j W_j \,\%\mathrm{HI}_{C,W_j}}{\sum_j W_j}, \tag{2}$$

where $\%\mathrm{HI}_{C,W_j}$ is the worst-case component index in group j (e.g. cables, joints) and W_j is the group weight. The worst-case rule conservatively ensures that a single severely degraded component is not masked by healthy peers.

B. Conditional Factor and Weibull Adjustment

Static indices ignore operating stress. To make the HI dynamic and reflective of real-world conditions (e.g. high loading, harsh environments), a *Conditional Factor* (CF) is introduced. The CF is a normalised, weighted score over operational covariates such as time since last maintenance, asset age, inspection/maintenance history, load, and operating environment. From the raw weighted score we obtain

$$\mathrm{CF} = \frac{\text{total weighted score}}{\text{total weightage}}, \qquad \%\mathrm{CF} = \frac{4 - \mathrm{CF}}{4} \times 100, \tag{3}$$

where the four-level scale follows the convention that rank 0 denotes excellent condition and rank 4 the worst condition; $\%\mathrm{CF}$ therefore increases with degradation. The CF modulates a Weibull survival function used to shape the degradation curve:

$$S(t) = \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right], \tag{4}$$

where t is time in years, α is the characteristic life (e.g. 40 years), and the shape parameter β is driven by the Conditional Factor:

$$\beta = 2 + \frac{\%\mathrm{CF}}{100} \times 8. \tag{5}$$

A higher $\%\mathrm{CF}$ thus steepens the survival curve, accelerating the modelled decline for assets operating under heavy stress.

C. Configuration-Driven ML Pipeline

The HI service decouples domain knowledge from code through a JSON configuration. Each parameter declares its name, an `operationType` (e.g. `upperLimit`, rate-of-change, or categorical `equality`), a Condition Assessment Scale (`cas`) mapping rank strings to numeric intervals or categorical lists, an optional *with-respect-to* reference, and a `weightage`. A representative entry is shown in Listing 1.

Algorithm 1 Offline HI model construction

1: Load JSON parameter configuration and filter inactive parameters
2: Parse and cache CAS rules per parameter
3: Simulate / ingest records; compute HI via Eq. (1)–(2)
4: Split into train/test (80/20)
5: Train LightGBM regressor; evaluate R^2, MAE
6: Serialise model (joblib) and export to ONNX


```
{
  "param_name": "gas_pressure_measured",
  "operationType": "upperLimit",
  "withRespectTo": "",
  "typeDataType": 0,
  "cas": {
    "0": "[0.0, 7.0]",
    "4": "[7.0, inf]"
  },
  "weightage": 25.0
}
```


Listing 1. Representative parameter configuration (illustrative; values generalised).

The expert-elicitation workflow proceeds as follows: (1) domain experts evaluate the significance of each inspection/testing criterion; (2) weightings are assigned out of 100 and documented for transparency and consistency; (3) JSON structures are created for both data simulation and HI computation; (4) synthetic parameter values are generated according to the CAS ranges and per-rank probabilities, optionally conditioned on a with-respect-to value; and (5) the HI is computed per record by dispatching each parameter to the operator implied by its `operationType` and accumulating weighted contributions. The scoring engine pre-parses and caches CAS rules, then evaluates records in parallel chunks across CPU cores for throughput.

The computed (features, HI) pairs constitute a supervised learning dataset. A LightGBM gradient-boosted regressor [16] is trained to predict HI from raw measurements, using an 80/20 train–test split and reporting the coefficient of determination R^2 and mean absolute error (MAE). Algorithm 1 summarises the offline pipeline.

D. Portable Inference Service

After training, the model is exported to the Open Neural Network Exchange (ONNX) format [17] for fast, framework-agnostic, cross-platform deployment. A lightweight REST service loads the ONNX model with an inference runtime and exposes a per-subsystem prediction endpoint. For each request the service: (i) validates the requested subsystem against a model map that records weightage, model file, and expected feature columns; (ii) preprocesses the incoming asset payload, filling any missing feature using its rank-0 CAS rule (categorical default or uniform draw within the rank-0 interval) and reordering columns to the model's training order; (iii) runs single-batch inference; and (iv) returns the predicted HI together with data-quality indicators. Two such indicators are reported:

$$\text{Data Availability } (\%) = \frac{\#\,\text{features provided}}{\#\,\text{features required}} \times 100, \tag{6}$$

$$\text{Coverage Rate } (\%) = \frac{\text{Data Availability} \times w_{\text{sub}}}{100}, \tag{7}$$

where w_{sub} is the subsystem weightage. These metrics let downstream consumers gauge how much trust to place in each prediction. A representative request/response pair is shown in Listing 2.

```
// Request: POST /AHI/<asset>/<subsystem>
{ "asset": { "gas_pressure_measured": 4.73,
             "sf6_purity": 99.3, "...": 0 } }

// Response
{ "Prediction": 0.0039,
  "Data Availability (%)": 100.0,
  "Coverage Rate (%)": 20.0,
  "status": true }
```

Listing 2. Representative HI inference request and response.

V. Remaining Life Assessment (RLA) Service

The RLA service estimates the residual operational life of an asset by fitting a degradation model to its historical HI trajectory. It currently supports motor and cable asset classes and integrates with the monitoring solution via a single REST endpoint. Because the historical system-HI series is supplied by the HI service, the RLA stage only needs to compute the Conditional Factor and fit the degradation curve. The layered RLA architecture is shown in Fig. 3.

A. Conditional Factor Scoring

The CF quantifies present health from operational covariates through weighted scoring, dispatched to asset-specific handlers. For a motor, covariates include time since last maintenance, machine age, past inspection/maintenance problems, repair history, current load, operating environment, OEM experience, and time since last lubrication. For a cable, covariates include commissioning age, load current, cumulative failures and repairs, network reliability (radial/loop/ network), length, laying environment, and whether PD faults have been identified. Each covariate is scored on the four-level scale; the normalised result feeds Eq. (3) to obtain $\%$CF, and Eq. (5) fixes the Weibull shape parameter β.

B. Hybrid Weibull–Polynomial Degradation Model

To capture both asset-specific trends and probabilistic ageing, the service fits a hybrid model that multiplies a polynomial trend by the Weibull survival term:

$$H(t) = \underbrace{P(t)}_{\text{polynomial}} \cdot \underbrace{\exp\Big[-\big(\tfrac{t}{\alpha}\big)^{\beta}\Big]}_{\text{Weibull}}, \tag{8}$$

where $P(t)$ is a low-order polynomial whose coefficients are obtained by curve-fitting to the (smoothed, rescaled) historical health series and α, β are fixed from the CF analysis. The polynomial captures local trends while the Weibull factor enforces a physically plausible long-term decline.

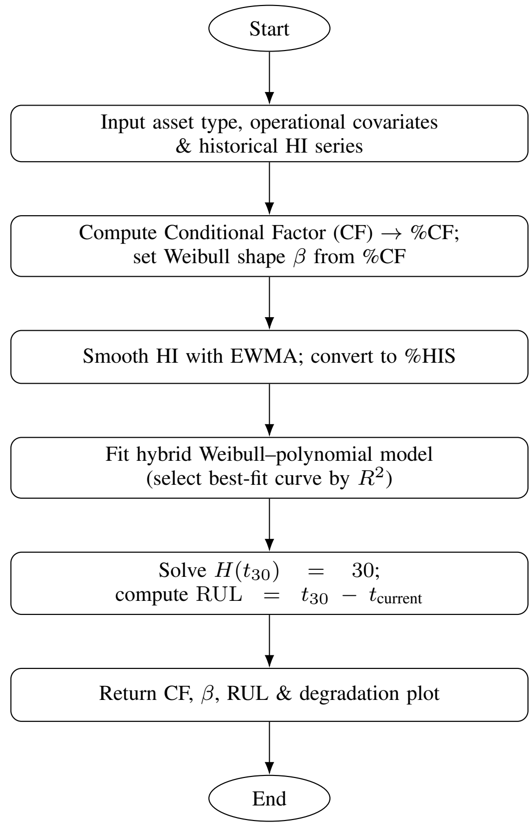

Fig. 3. Processing pipeline of the Remaining Life Assessment (RLA) service.

C. Smoothing and Scale Conversion

Raw HI readings are noisy. The service applies an exponentially weighted moving average (EWMA),

$$\hat{H}_t = a\,H_t + (1-a)\,\hat{H}_{t-1}, \tag{9}$$

with the smoothing factor a selected automatically to balance responsiveness and noise rejection. The smoothed score is converted to a percentage remaining-health scale ($\%$HIS): on the four-level rank (0 excellent to 4 worst), a pristine asset starts near 100% and decays toward 0% as it deteriorates,

$$\%\text{HIS} = \frac{4-\hat{H}_t}{4} \times 100, \tag{10}$$

so that higher values denote a healthier asset and the end-of-life threshold is a low $\%$HIS. Smoothing reduces measurement noise, sharpens trend detection, improves model fit, and suppresses false alarms from transient spikes.

D. RUL by Threshold Crossing

The Remaining Useful Life is the time until the projected health curve falls to an end-of-life threshold (e.g. $\%\text{HIS} = 30$). Solving

$$H(t_{30}) = 30 \tag{11}$$

numerically (e.g. with a nonlinear root finder) yields the crossing time t_{30}, and the remaining life is

$$\text{RUL} = t_{30} - t_{\text{current}}. \tag{12}$$

Algorithm 2 Remaining Life Assessment

Receive asset type, covariates, and historical HI series
Compute CF; derive %CF (Eq. (3)) and β (Eq. (5))
Smooth HI with EWMA (Eq. (9)); convert to %HIS (Eq. (10))
Fit hybrid Weibull–polynomial model (Eq. (8))
Solve $H(t_{30}) = 30$ (Eq. (11)); compute RUL (Eq. (12))
Return CF, β, RUL, and degradation plot

The service returns the Conditional Factor, the fixed β, the RUL estimate, and a degradation plot annotating the current state and the projected threshold crossing. Algorithm 2 summarises the procedure.

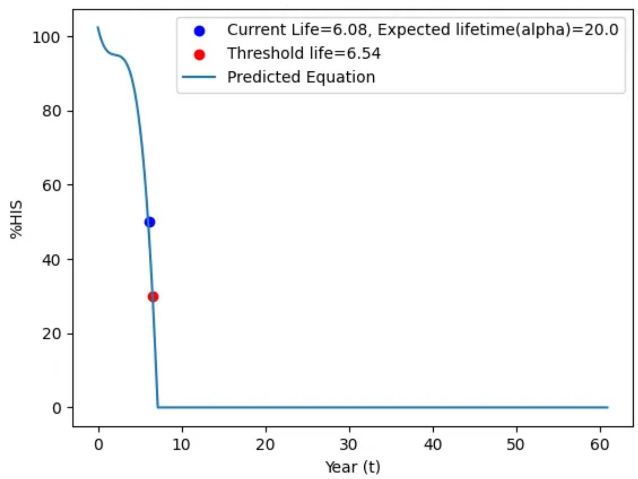

Fig. 4. Residual Life Assessment for a motor nearing end of life. The fitted Weibull–polynomial model projects %HIS over time; the current state (%HIS $\approx$ 50% at $t = 6.08$ yr) sits just above the 30% end-of-life threshold ($t = 6.54$ yr), yielding a remaining useful life of only ≈ 0.46 yr against a characteristic lifetime $\alpha = 20$ yr.

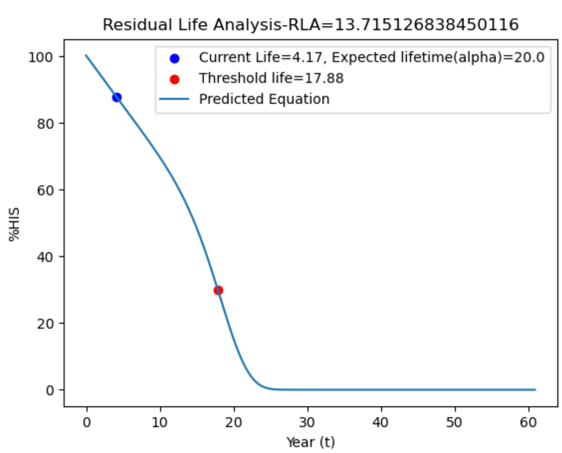

Fig. 5. Residual Life Assessment for an asset in good condition. The current state (%HIS $\approx$ 88% at $t = 4.17$ yr) lies well above the 30% end-of-life threshold ($t = 17.88$ yr), giving a substantial remaining useful life of ≈ 13.72 yr against a characteristic lifetime $\alpha = 20$ yr.

VI. Probability of Failure (PoF) Service

Whereas RLA returns a point (or interval) estimate of remaining life, the PoF service quantifies the likelihood that an asset fails within a finite horizon (e.g. 7, 30, or 90 days), combining HI and RLA outputs. The proposed formulation rests on survival analysis. Let T denote the time to failure with survival function $S(t) = \Pr(T > t)$ and hazard $h(t)$. The probability of failure within a horizon $[t, t + \Delta]$ conditional on survival to t is

$$\mathrm{PoF}(t, \Delta) = \frac{S(t) - S(t + \Delta)}{S(t)} = 1 - \exp\left[-\int_{t}^{t+\Delta} h(u)\, du\right]. \tag{13}$$

The hazard can be learned from historical HI trajectories and recorded failure events using a Cox proportional hazards model [11] or a random survival forest [12], with HI level, HI rate-of-change, CF, and environmental covariates as predictors. Because the inputs (HI forecasts) are themselves uncertain, Monte Carlo simulation propagates that uncertainty: repeated sampling of plausible HI trajectories yields a distribution of hazard/survival realisations, from which mean PoF and confidence bands are obtained. Outputs include per-horizon PoF metrics, fleet-level risk rankings, and heatmaps that integrate with maintenance-planning and ERP systems.

VII. Partial Discharge Monitoring Subsystem

Partial discharge (PD) is a localised breakdown that only partially bridges the insulation between conductors, and it is a leading indicator of insulation degradation in high-voltage assets [13], [14]. The HI service consumes PD severity as a diagnostic input; this section describes the supporting acquisition and phase-resolved partial-discharge (PRPD) handling.

A. Acquisition and Storage

High-frequency PD sensors feed dedicated analog-to-digital converter channels whose sampling rates resolve nanosecond-scale discharges, preserving amplitude and timing. Digitised data are encapsulated in a custom transport protocol carrying synchronisation markers, per-channel metadata, and error-checking codes for integrity in electrically noisy environments. Every five minutes the acquisition gateway computes summary metrics—maximum amplitude (PD_MaxAmp), maximum discharge rate (PD_MaxRate), average amplitude (PD_AvgAmp), and average discharge rate (PD_AvgRate)—and derives output indicators including a Boolean PD status and a composite severity level. Metrics are serialised to JSON and archived per asset and timestamp.

B. Time-Bucketed File Convention

Each archive file follows the pattern `005mins_S$_YYYY_MM_DD_hh_mm.txt`, where `005mins` denotes the five-minute interval, `S0`/`S1` flag the absence/presence of PD, and the timestamp uses minute multiples of five. This convention enables simple pattern queries (e.g. all PD-positive files on a date) and chronological sorting. Amplitudes are expressed in millivolts (mV) and discharge in picocoulombs (pC) to prevent conversion errors.

C. PRPD Representation and Visualisation

PRPD records are stored either as aggregated summaries or as full-scale waveform segments encoded as delimited strings (<PhaseAngle>!<Amplitude>!<Count> or <Amplitude>!<Count>, with segments separated by @). A decoupled pipeline—capture, frame and transport, gateway processing, message queueing, and ingestion into snapshot and history tables—supports high throughput and fault tolerance. A visualisation API selects PD-positive files, groups them by year/month/day/hour/minute, and merges consecutive five-minute files for wider time windows by combining counts and normalising amplitudes before plotting the PRPD pattern. Configurable retention and purge policies bound storage growth.

VIII. Discussion

The framework offers several practical advantages. *Flexible configuration* through JSON-defined parameters and a subsystem-to-model map allows new assets and models to be onboarded without changing core code. *Robust preprocessing* automatically fills missing features using rank-0 CAS rules, computes data-availability metrics, and enforces correct feature ordering. *High-performance inference* via ONNX runtime supports both batch and near-real-time use. *Comprehensive metrics*—returning not only the HI but also data-availability and coverage—empower downstream systems to gauge prediction trust.

Several limitations and threats to validity remain. The synthetic data generation used to bootstrap the HI model encodes expert assumptions through per-rank probabilities; if these diverge from field distributions, the learned mapping may be biased, motivating periodic retraining on real measurements with data-drift detection. The hybrid degradation model assumes a monotone long-term decline modulated by a fixed Weibull shape; assets exhibiting self-healing or step changes after maintenance may require piecewise or Bayesian-updated models. Finally, the PoF service depends on a sufficient history of labelled failure events, which is often scarce for highly reliable assets; censored-data techniques and fleet-level pooling mitigate but do not eliminate this constraint.

IX. Conclusion and Future Work

We presented an integrated, microservice-based framework that unifies a dynamic Health Index service, a Remaining Life Assessment service, and a Probability of Failure service for critical electrical assets, supported by a partial-discharge monitoring subsystem. The HI service combines expert-weighted scoring, condition-based adjustment via a Conditional Factor, and a Weibull survival formulation, all learned by a gradient-boosted regressor and served through portable ONNX inference. The RLA service projects remaining life from smoothed HI trajectories using a hybrid Weibull–polynomial model, and the PoF service quantifies finite-horizon failure risk through survival analysis and Monte Carlo uncertainty propagation.

Future work includes: (i) multi-sensor fusion of HI with vibration, acoustic, and thermal indicators; (ii) temporal feature engineering (rolling statistics, volatility, rate-of-change); (iii) automated retraining pipelines with data-drift detection using MLOps tooling; (iv) explainability via SHAP/LIME and actionable risk-mitigation recommendations; and (v) containerised orchestration for elastic scaling across batch and low-latency online modes. Together these enhancements evolve the platform from a static HI predictor into a full asset-reliability suite delivering prognostics, risk quantification, and prescriptive maintenance guidance.

References

[1] A. Jahromi, R. Piercy, S. Cress, J. Service, and W. Fan, “An approach to power transformer asset management using health index,” *IEEE Electrical Insulation Magazine*, vol. 25, no. 2, pp. 20–34, 2009.

[2] A. Naderian, S. Cress, R. Piercy, F. Wang, and J. Service, “An approach to determine the health index of power transformers,” in *Proc. IEEE Int. Symp. Electrical Insulation (ISEI)*, 2008, pp. 192–196.

[3] M. M. Islam, G. Lee, and S. N. Hettiwatte, “Determining health index of transmission line asset using a condition-based method,” *Resources*, vol. 8, no. 2, art. 80, 2019.

[4] Research on estimating the insulation health index of XLPE cables using machine learning, *Scientific Reports / PMC*, 2024.

[5] “Health index for power transformer condition assessment based on operation history and test data,” *Energy Reports*, vol. 8, pp. 9038–9050, 2022.

[6] W. A. Tamma, R. A. Prasojo, and Suwarno, “High voltage power transformer condition assessment considering the health index value and its decreasing rate,” *High Voltage (IET)*, vol. 6, no. 2, pp. 314–327, 2021.

[7] “Power transformer health index and life span assessment using machine learning,” *Engineering Applications of Artificial Intelligence*, 2024.

[8] W. Weibull, “A statistical distribution function of wide applicability,” *Journal of Applied Mechanics*, vol. 18, no. 3, pp. 293–297, 1951.

[9] J. B. Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, and F. Fnaiech, “Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network,” *Mechanical Systems and Signal Processing*, vol. 56–57, pp. 150–172, 2015.

[10] “A comprehensive review of remaining useful life estimation approaches for rotating machinery,” *Energies*, vol. 17, no. 22, art. 5538, 2024.

[11] D. R. Cox, “Regression models and life-tables,” *Journal of the Royal Statistical Society: Series B*, vol. 34, no. 2, pp. 187–220, 1972.

[12] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer, “Random survival forests,” *The Annals of Applied Statistics*, vol. 2, no. 3, pp. 841–860, 2008.

[13] “Review on partial discharge diagnostic techniques for high voltage equipment in power systems,” *IEEE Access*, vol. 11, 2023.

[14] IEC 60270, *High-Voltage Test Techniques—Partial Discharge Measurements*, International Electrotechnical Commission, 2015.

[15] IEEE Std 400.3-2022, *IEEE Guide for Partial Discharge Field Diagnostic Testing of Shielded Power Cable Systems*, IEEE, 2022.

[16] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, “LightGBM: A highly efficient gradient boosting decision tree,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 3146–3154.

[17] Open Neural Network Exchange (ONNX). [Online]. Available: https://onnx.ai