CT X-ray Tube RUL Estimation Under Data Scarcity in Low-Resource Hospitals

Zakaria Oubnini and Feng Wang

Henan University of Technology, Zhengzhou, China
oubniniz@gmail.com

Abstract. CT X-ray tubes are among the most expensive consumables in medical imaging, and in low-resource hospitals each unplanned failure causes weeks of scanner downtime. Predictive maintenance through Remaining Useful Life (RUL) estimation could prevent this, but real hospital logs record only sparse failure events, not the continuous sensor data deep models need. We calibrate a physics-based degradation simulator to real failure records from a Moroccan public hospital (CHU Ibn Rochd), reproducing the observed mean tube lifespan within 4.0%, and use it to generate synthetic RUL-labelled trajectories. On these we benchmark six models, from classical regressors to a physics-informed LSTM (PINN-LSTM), under progressively scarcer training data (100%, 50%, 20%) over five seeds. Two findings emerge. First, model simplicity drives robustness: as data shrinks to 20%, lightweight recurrent models stay below 3.8 days RMSE while CNN-LSTM and Transformer architectures exceed 5.4 days. Second, the physics-informed monotonicity constraint matches the plain LSTM's accuracy and improves physical validity only when data is abundant, offering no advantage under scarcity; a C-MAPSS replication confirms its effect is small and dataset-dependent. In data-scarce clinical settings, simpler models are more reliable than added architectural or physics-based complexity.

Keywords: Remaining Useful Life · CT X-ray Tubes · Data Scarcity

1 Introduction

Computed tomography (CT) is central to modern hospital diagnosis, yet access in Morocco's public health system is severely constrained: the public sector operates only 176 CT scanners for 37.5 million people [1], about 4.7 per million, roughly a quarter of France's 20 per million [2]. The fleet is already overloaded, with CHU Mohammed VI (Marrakech) recording 8,048 examinations per scanner per year [4] and CT wait times at CHU Ibn Sina (Rabat) rising 61% in a single year [3].

Scanner breakdowns compound this scarcity. A national audit by Morocco's Cour des Comptes found the Ministry of Health had stopped tracking equipment breakdown history after 2011, with its maintenance management system deployed to 87 delegations but used by only 12 [7]; repairs at CHU Ibn Rochd

have taken up to four days per incident [5], and the hospital's strategic plan flagged the absence of such a system [8]. Economically, tube replacement costs 50,000 DH and annual contracts 895,000 DH (about 20% of acquisition cost) [7], while private-sector exams cost roughly 140 EUR [6]. A low-cost predictive-maintenance capability would therefore be especially valuable.

The X-ray tube is the principal source of unplanned CT downtime. Its degradation follows physical processes, tungsten evaporation, filament burnout, and anode thermal stress [9], that render Remaining Useful Life (RUL) estimation tractable in principle, and recurrent deep models such as the LSTM dominate equipment RUL estimation [16] and have begun to be applied to CT tubes [14, 15]. Two critical gaps remain: existing models do not enforce physical degradation constraints, allowing physically impossible predictions of increasing RUL, and all published approaches assume large training datasets unavailable where hospital logs are sparse.

This paper investigates both gaps empirically. We calibrate a physics-based degradation simulator against failure records from CHU Ibn Rochd Casablanca ($\sim$300 examinations/day), reproducing its mean lifespan within 4.0%, and check it without re-fitting against a lower-workload hospital, Hospital Civil Tétouan ($\sim$12/day)[1]; we then use 100 synthetic trajectories from it to benchmark six models, from classical regressors to a physics-informed LSTM (PINN-LSTM) with a pair-wise monotonicity loss, under training-data fractions of 100%, 50%, and 20%. Our findings are cautionary: model simplicity, not complexity, governs robustness to data scarcity, while the physics-informed constraint improves prediction validity only when data is abundant. To our knowledge, this is the first systematic evaluation of model complexity and physics-informed constraints for CT tube RUL under data scarcity in low- and middle-income settings.

2 Related Work

Most prior machine-learning work on CT tube health frames the problem as fault *classification* rather than continuous prediction. Wang et al. [10] detected arcing from IoMT sensor data with a Naïve Bayes classifier, and Zhou et al. [11] improved on it with a symbolic-aggregate-approximation classifier identifying oil temperature as the most discriminative feature; the current state of the art is the Gated Fusion Network of Tang et al. [12], and Pomšár et al. [9] reached 87% accuracy with full recall on faulty tubes using LSTM and 1D-CNN models. All output a binary or categorical signal rather than a continuous time-to-failure estimate. Continuous RUL estimation for CT tubes has received far less attention: Shamim [14] applied a standard LSTM ($\pm$5% RUL error), Chen et al. [15] proposed DDLNet (MSE = 2.92 days^2 on six complete-lifecycle tubes), and Luan et al. [13] forecast failure *counts* but no per-tube RUL. Neither Shamim nor Chen enforces physical degradation constraints or evaluates performance under data scarcity, and both rely on datasets unavailable outside their own institutions.

[1] Maintenance logs obtained from a biomedical maintenance technician at Hospital Civil Tétouan (personal communication, 2026).

Physics-informed neural networks add physical-constraint terms to the training loss to enforce known degradation laws. Lu et al. [16] embedded a monotonic-degradation consistency penalty in an LSTM and improved bearing RUL accuracy, but no prior work has applied this idea to CT tubes. Table 1 summarises these differences: prior CT tube work is almost exclusively classification on proprietary data, without physical constraints or a deployment focus, whereas ours combines continuous RUL estimation, a physics-informed model, real-plus-synthetic data, and an explicit low-resource deployment context.

Table 1. Comparison of CT tube predictive maintenance approaches. Task: RUL = continuous remaining useful life, Cls = classification, Count = failure-count forecasting; Physics = physical degradation laws embedded in the model; Context = whether the study targets a specific real-world deployment setting (here, low-resource hospitals).

Reference	Task	Physics	Dataset	Context
Wang et al. [10]	Cls	No	Proprietary	No
Zhou et al. [11]	Cls	No	Proprietary	No
Tang et al. [12]	Cls	No	Proprietary	No
Luan et al. [13]	Count	No	Proprietary	No
Pomšár et al. [9]	Cls	No	Proprietary	No
Shamim [14]	RUL	No	Synthetic	No
Chen et al. [15]	RUL	No	Proprietary	No
This work	**RUL**	**Yes**	**Synthetic+real**	**Yes**

3 Methodology

Figure 1 gives an overview of the proposed framework. Because no public dataset of CT X-ray tube degradation exists, and real hospital logs record only sparse failure events rather than continuous sensor streams, we generate training data from a physics-based degradation simulator calibrated against real failure records from CHU Ibn Rochd, with a second, lower-workload hospital (Tétouan) used only as a held-out check. The simulator couples four mechanistic degradation processes, namely anode thermal behaviour, filament evaporation, thermal stress, and bearing wear, to produce 100 synthetic tube trajectories, each with nine measurable sensor channels and a ground-truth remaining useful life (RUL) label at every cycle. A physics-informed LSTM (PINN-LSTM) is then trained on these trajectories under a hybrid loss whose physics term enforces that predicted RUL can only decrease over time.

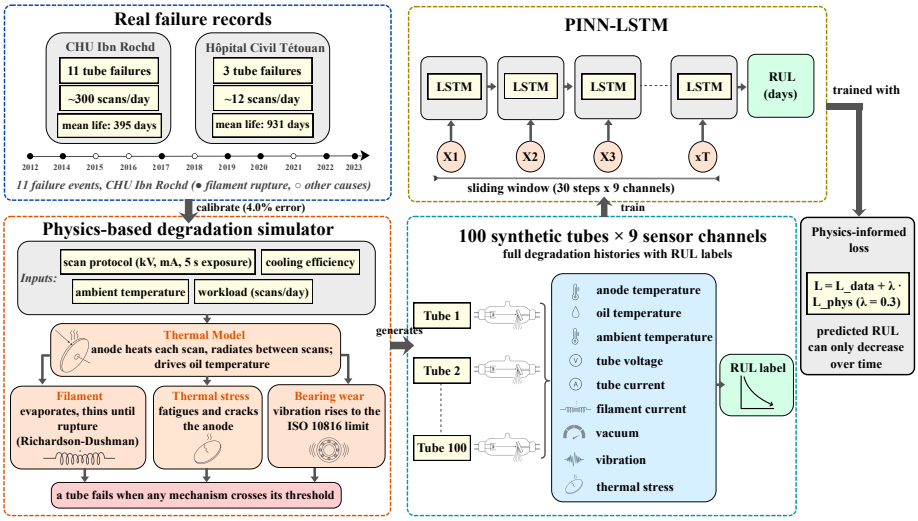

Fig. 1. Proposed framework. Real failure records calibrate a physics-based simulator; its synthetic tube trajectories train a PINN-LSTM under a monotonicity-enforcing loss.

3.1 Physics-Based Degradation Simulator

The simulator advances a single tube through repeated scan cycles, updating four coupled degradation processes (thermal, filament, thermal-stress, and bearing) and tracking vacuum pressure at each cycle. Each scan deposits heat $Q = VI\tau$ into the tungsten anode (tube voltage V, current I, exposure τ); between scans the anode cools radiatively, modelled as a lumped thermal mass

$$\rho c_p V_{\text{th}} \frac{dT}{dt} = Q - \sigma\varepsilon A\,(T^4 - T_{\text{amb}}^4), \tag{1}$$

where ρ, c_p, and V_{th} are the density, specific heat, and effective thermal volume of the tungsten anode; T and T_{amb} are the anode and ambient temperatures; t is time; Q is the heat deposited per scan; σ is the Stefan–Boltzmann constant; ε is the anode emissivity; and A is its radiating area. The radiative term is scaled by a 0.7 cooling-efficiency factor that reflects the degraded air-cooling and 32°C ambient documented at CHU Ibn Rochd. A coupled node tracks oil-jacket temperature, the strongest empirical predictor of tube faults [11]. The remaining processes follow standard degradation physics: filament emission follows the Richardson–Dushman law $J = A_R T^2 e^{-W/kT}$ (emission current density J, Richardson constant A_R, work function W, Boltzmann constant k), with the filament thinning by evaporation until rupture; thermal stress $\sigma_{\text{th}} = E\alpha(T - T_0)$ accumulates as fatigue (Young's modulus E, thermal expansion coefficient α, reference temperature T_0); and bearing wear raises rotor vibration toward the ISO 10816 limit. A tube reaches end-of-life when any process crosses its threshold. We calibrate the evaporation rate so that the simulated mean lifespan matches the CHU Ibn Rochd record: under the documented

regime (∼300 scans/day, 5 s exposure) the simulator yields a mean lifespan of 379 days against the observed 395, a 4.0% calibration error, with filament failure dominant as in the real data. As an independent check, we apply the same simulator, without re-fitting, to the much lower-workload Tétouan regime (∼12 scans/day): it over-predicts the observed 931-day lifespan by more than an order of magnitude, a limitation we examine in Section 5.1.

3.2 Synthetic Dataset and Features

The simulator advances each tube cycle by cycle until a degradation threshold is crossed, producing 100 synthetic tube histories. Each history is a multivariate time series of nine sensor channels sampled per scan cycle, with a ground-truth RUL label, the number of cycles remaining until failure (capped at 50,000), at every step. The nine channels span temperatures, electrical signals, and condition indicators, all measurable on a deployed scanner (Fig. 1). Internal states such as filament diameter and cumulative stress are excluded as inputs, since they cannot be observed and would leak the degradation state. The real records are used only to calibrate and check the simulator, not for training, because they contain only a handful of failure events, each a date and a failure cause (for example filament rupture versus other causes, as in Fig. 1), rather than the continuous per-cycle sensor trajectories that supervised learning requires: CHU Ibn Rochd contributes 11 replacements (2012–2023; mean 395 days at ∼300 scans/day) and Hospital Civil Tétouan 3 (mean 931 days at ∼12 scans/day), the latter held out as an independent check. Each history is segmented into overlapping windows of 30 cycles, each labelled with the RUL at its final step, and channels are min–max normalised. Tubes are split 70/15/15 into training, validation, and test sets, fixed across all runs; to probe data scarcity, the training tubes are subsampled to 100%, 50%, and 20%. The validation set is used only for early stopping and model selection; all reported results, including Tables 2 and 3, are computed on the held-out test set, which is never seen during training.

3.3 Physics-Informed LSTM

The predictor is a stacked LSTM (two layers, 64 hidden units) that maps each 30-step, nine-channel window to a normalised RUL through a fully connected head. Training minimises a hybrid objective

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda\, \mathcal{L}_{\text{phys}}, \qquad \lambda = 0.3, \tag{2}$$

where $\mathcal{L}_{\text{data}}$ is the mean squared error between predicted and true RUL. The physics term enforces the one constraint every degrading tube must obey: remaining life cannot increase. For two windows drawn from the same tube at times t and $t{+}k$, with predictions $\hat{y}_{\text{early}}$ and $\hat{y}_{\text{late}}$, we penalise any violation of monotonic decrease,

$$\mathcal{L}_{\text{phys}} = E\Big[\, \big(\text{ReLU}(\hat{y}_{\text{late}} - \hat{y}_{\text{early}})\big)^2 \,\Big]. \tag{3}$$

For example, a later window predicted to have less remaining life than an earlier one ($\hat{y}_{\text{late}} = 100 < \hat{y}_{\text{early}} = 150$) gives a negative difference, so ReLU returns zero and no penalty applies, as desired; a predicted increase ($\hat{y}_{\text{late}} = 150 > \hat{y}_{\text{early}} = 100$) gives a positive difference and a quadratic penalty, discouraging the physically impossible case. A standard LSTM trained on shuffled windows never observes the relationship between successive states of the same tube and can predict physically impossible RUL increases; the pair-wise term removes this freedom. All models use the Adam optimiser (learning rate 10^{-3}), early stopping on validation loss, and fixed random seeds; the same architecture with $\lambda = 0$ serves as the baseline LSTM.

4 Experiments and Results

4.1 Experimental Setup

We compare PINN-LSTM against five baselines: support-vector regression (SVR), random forest (RF), a plain LSTM, a CNN-LSTM, and a Transformer encoder. All deep models share the same two-layer, 64-unit backbone; PINN-LSTM differs from the plain LSTM only in the added monotonicity loss ($\lambda = 0.3$). The classical baselines are fitted on flattened windows, subsampled to 30,000 windows for tractability. We report RMSE in days (one day = 300 scan cycles; e.g., the scarcest LSTM error of 3.50 days is 1050 cycles) as the primary metric; MAE and R^2, omitted for space, follow the same trends. To assess robustness to data scarcity, each model is trained on 100%, 50%, and 20% of the training tubes while the validation and test sets are held fixed; reported values are means over five random seeds.

Data use declaration. The real-world data used to calibrate the simulator consist solely of anonymized CT-tube failure dates and causes, obtained from hospital maintenance records and publicly available activity and audit reports; they contain no patient information, and no ethics approval was therefore required. The C-MAPSS turbofan dataset [17] is publicly released by NASA. The synthetic degradation trajectories used for training and evaluation are produced by the simulator described in Section 3.1.

4.2 Results under Data Scarcity

Table 2 reports RMSE for all six models at each training fraction. At full data the deep models are closely matched (2.6 to 3.1 days), with CNN-LSTM marginally best and the physics constraint neither helping nor hurting accuracy. The decisive effect appears as data shrinks. Between 100% and 20% of the training tubes the complex architectures collapse, CNN-LSTM by 109% (2.61 to 5.45 days) and the Transformer by 82% (3.14 to 5.71), while the two lightweight recurrent models degrade far less, the plain LSTM by 27% (2.76 to 3.50) and PINN-LSTM by 39% (2.71 to 3.77). At the scarcest setting the plain LSTM is the most accurate

model (3.50 days), with PINN-LSTM close behind (3.77) and statistically indistinguishable given the seed variance; both decisively outperform CNN-LSTM (5.45) and the Transformer (5.71), and the classical baselines fare worst (random forest 9.94, SVR 43.71 days). The central finding concerns model complexity rather than physics: simple recurrent models are markedly more robust to data scarcity than deeper architectures, and the physics-informed constraint yields no accuracy advantage over a plain LSTM.

Table 2. RMSE in days (mean ± std over five seeds) by model and training-data fraction. Lower is better; the best value in each column is in bold.

Model	100%	50%	20%
SVR	35.07 ±0.13	33.46 ±0.08	43.71 ±0.00
Random Forest	4.60 ±0.08	5.45 ±0.05	9.94 ±0.10
LSTM	2.76 ±0.13	**2.80 ±0.06**	**3.50 ±0.46**
CNN-LSTM	**2.61 ±0.16**	2.94 ±0.21	5.45 ±0.52
Transformer	3.14 ±0.39	3.03 ±0.16	5.71 ±0.60
PINN-LSTM (ours)	2.71 ±0.18	2.94 ±0.24	3.77 ±0.26

4.3 Physical Consistency of Predictions

The physics constraint is designed not to improve accuracy but to keep predictions physically valid: a tube's remaining life cannot increase over time. Table 3 reports the fraction of same-tube window pairs for which the predicted RUL increases, a physically impossible event, at separations of 5, 10, and 20 windows. At full data the constraint behaves as intended, with PINN-LSTM roughly halving long-range violations relative to the plain LSTM (10.1% versus 17.1% at a 20-window separation). Under data scarcity, however, this benefit vanishes: the two models become indistinguishable (18.1% versus 17.4% at the same separation). Physical consistency, like accuracy, improves only when data is sufficient. In absolute terms the violation rates stay substantial: even PINN-LSTM's best case leaves 10.1% of 20-window pairs violating monotonicity at full data, rising to roughly 18% under scarcity, so neither model is close to fully consistent.

Table 3. Monotonicity violations (% of same-tube window pairs where predicted RUL increases; lower is better).

	Full data (100%)		Scarce data (20%)	
Pair separation	LSTM	PINN-LSTM	LSTM	PINN-LSTM
5 windows	37.1	**33.3**	**37.4**	38.5
10 windows	28.2	**22.4**	**28.8**	30.0
20 windows	17.1	**10.1**	**17.4**	18.1

4.4 Replication on the C-MAPSS Benchmark

To test whether the physics constraint generalises beyond our simulator, we replicate the LSTM-versus-PINN comparison on the public NASA C-MAPSS benchmark [17]. The absolute test RMSE (cycles) at training fractions of 10/25/50/100% is 13.71/13.14/13.08/13.02 for the plain LSTM and 13.67/12.87/12.84/13.36 for PINN-LSTM: the constraint lowers RMSE by 2.0% and 1.9% at 25% and 50% but is marginally worse (−2.6%) at full data, the opposite of the CT pattern. Across both datasets the physics constraint yields only small, dataset-dependent effects, whereas the robustness advantage of simple over complex architectures is large and consistent across every data fraction in our CT experiments.

5 Discussion

Our results carry a different message from the one we set out to demonstrate: the dominant effect is model capacity, not physics. As the training set shrinks, the high-capacity architectures (CNN-LSTM, Transformer) overfit and collapse, while lightweight recurrent models retain most of their accuracy, a concrete bias-variance trade-off in which extra architectural machinery becomes a liability under scarce labels. The physics-informed constraint proved neither harmful nor decisive: it matched the plain LSTM's accuracy at every level and, at full data, roughly halved long-range monotonicity violations, but under scarcity both benefits vanished and its small C-MAPSS gain went the opposite way. A soft pair-wise penalty is too weak to shape predictions when labels are themselves scarce. The practical lesson: resource-constrained departments keep only sparse failure logs, so the safe choice is a simple recurrent model, which degrades gracefully where complex or physics-augmented ones fail, at a fraction of the cost.

5.1 Limitations

Several limitations temper these conclusions. First, results are on synthetic trajectories: the simulator is calibrated to real records but not validated on real CT sensor streams (unavailable to us), so this is a controlled study, not field validation. Second, calibration is anchored to one high-workload hospital (14 real tubes, filament-dominated); applied without re-fitting to the low-workload Tétouan regime it over-predicts lifespan by over an order of magnitude, indicating that at low throughput, common in low-resource settings, failures are driven by calendar-age effects (vacuum degradation, oxidation) the usage-driven model omits, so age-aware, multi-site calibration is essential. Third, we test a single soft constraint at one weight; a stronger formulation and a weight-sensitivity analysis would sharpen the conclusions. Finally, the efficiency argument rests on architecture rather than measured cost; FLOPs, latency, memory, and real-sensor validation across sites are the key next steps.

6 Conclusion

We studied CT X-ray tube RUL estimation under data scarcity, using a physics-based simulator calibrated to real failure records from CHU Ibn Rochd to benchmark six models across decreasing training data. The clearest finding is that model simplicity, not added complexity or physics, governs robustness: as data falls to 20%, CNN-LSTM and Transformer architectures degrade by 82 to 109% while lightweight recurrent models degrade by only 27 to 39%, and the physics-informed constraint helps only when data is abundant. We therefore recommend simple, robust models for resource-constrained settings, and caution that soft physics-informed regularisation is not a substitute for data; future work will pursue real CT sensor validation, multi-site calibration, and stronger constraints.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ministère de la Santé et de la Protection Sociale: Santé en Chiffres 2023. Direction de la Planification et des Ressources Financières, Rabat, Maroc (2023)
2. OECD: Health at a Glance 2023: OECD Indicators. OECD Publishing, Paris (2023). https://doi.org/10.1787/7a7afb35-en
3. Centre Hospitalo-Universitaire Ibn Sina: Rapport d'Activités 2023. Direction du Centre Hospitalo-Universitaire Ibn Sina, Rabat, Maroc (2023)
4. Centre Hospitalo-Universitaire Mohammed VI de Marrakech: Rapport d'Activité 2020–2021–2022. Division de l'Organisation et du Suivi des Affaires Professionnelles, Marrakech, Maroc (2022)
5. El Oualidi, M.A., Mousrij, A., Hajri-Gabouj, S., Darmoul, S.: Modélisation et analyse du flux des patients au service des urgences – CHU Ibn Rochd, Casablanca. In: Proceedings of the IXe Conférence Internationale Conception et Production Intégrées (CPI 2009), Fès, Maroc (2009)
6. Radi, S.: Interactions hospitalières au Maroc: les politiques d'accès aux soins dans la perspective des patients. Mondes en Développement **47**(187), 71–84 (2019).
7. Cour des Comptes du Royaume du Maroc: Rapport sur la gestion du parc des équipements biomédicaux dans les hôpitaux publics. Bulletin Officiel no. 6344-bis, pp. 1159–1192. Rabat, Maroc (2015).
8. Centre Hospitalier Universitaire Ibn Rochd: Projet d'Établissement Hospitalier 2019–2023. Direction du Centre Hospitalier Universitaire Ibn Rochd, Casablanca, Maroc (2019)
9. Pomšár, L., Tsvietaieva, M., Krupáš, M., Zolotova, I.: Classifying X-Ray tube malfunctions: AI-powered CT predictive maintenance system. Applied Sciences **15**(12), 6547 (2025). https://doi.org/10.3390/app15126547
10. Wang, C., Liu, Q., Zhou, H., et al.: Anomaly prediction of CT equipment based on IoMT data. BMC Medical Imaging **23**, 166 (2023). https://doi.org/10.1186/s12880-023-01037-0
11. Zhou, H., Liu, Q., Liu, H., Chen, Z., Huang, J.: Healthcare facilities management: a novel data-driven model for predictive maintenance of computed tomography equipment. Artificial Intelligence in Medicine **149**, 102807 (2024). https://doi.org/10.1016/j.artmed.2024.102807

12. Tang, Y., Zhou, Y., Wu, T., Wang, C., Li, Z., Li, K.: AI-driven predictive maintenance for medical imaging equipment: a deep learning framework based on the IoMT data. Reliability Engineering and System Safety **270**, 112152 (2026). https://doi.org/10.1016/j.ress.2025.112152
13. Luan, X., Yu, S., Yin, S., Liu, Z., Wang, D., Xiao, Y., Pan, R., Jia, B., Xu, F., Tian, Y., Li, R.: Temporal trends and reliability of computed tomography scanner failures: evidence from a real-world time-series analysis. Frontiers in Physics **14**, 1771484 (2026). https://doi.org/10.3389/fphy.2026.1771484
14. Shamim, M.M.R.: AI-driven predictive maintenance for high-voltage X-ray CT tubes: a manufacturing perspective. Review of Applied Science and Technology **3**, 40–67 (2024). https://doi.org/10.63125/npwqxp02
15. Chen, Z., Liu, Y., Qin, Z., Li, H., Xie, S., Fan, L., Liu, Q., Huang, J.: Dual-branch deep learning with dynamic stage detection for CT tube life prediction. Sensors **25**(15), 4790 (2025). https://doi.org/10.3390/s25154790
16. Lu, W., Wang, Y., Zhang, M., Gu, J.: Physics guided neural network: remaining useful life prediction of rolling bearings using long short-term memory network through dynamic weighting of degradation process. Eng. Appl. Artif. Intell. **127**, 107350 (2024). https://doi.org/10.1016/j.engappai.2023.107350
17. Saxena, A., Goebel, K., Simon, D., Eklund, N.: Damage propagation modeling for aircraft engine run-to-failure simulation. In: 2008 International Conference on Prognostics and Health Management, pp. 1–9. IEEE (2008). https://doi.org/10.1109/PHM.2008.4711414