Machine Learning for Covariate-Driven Changes in Weibull Scale Parameter

Muhammad Fitran Ramadhan
Computer Engineering Study Program, School of Electrical Engineering, Telkom University, Main Campus (Bandung Campus), Bandung, West Java, Indonesia
fitranhp@gmail.com

Nabil De Nozyra
Computer Engineering Study Program, School of Electrical Engineering, Telkom University, Main Campus (Bandung Campus), Bandung, West Java, Indonesia
nabil.denozyra@gmail.com

Fahrial Al Akbar Warsito Putra
Computer Engineering Study Program, School of Electrical Engineering, Telkom University, Main Campus (Bandung Campus), Bandung, West Java, Indonesia
fahrialalakbarwp@gmail.com

Surya Michrandi Nasution
Computer Engineering Study Program, School of Electrical Engineering, Telkom University, Main Campus (Bandung Campus), Bandung, West Java, Indonesia
michrandi@telkomuniversity.ac.id (corresponding author)

Reza Rendian Septiawan
Computer Engineering Study Program, School of Electrical Engineering, Telkom University, Main Campus (Bandung Campus), Bandung, West Java, Indonesia
zaseptiawan@telkomuniveristy.ac.id

Received: 28 September 2025 | Revised: 19 November 2025 | Accepted: 25 November 2025

 DOI: https://doi.org/10.48084/etasr.15186

ABSTRACT

This research addresses the limitations of conventional, static survival models for Remaining Useful Life (RUL) estimation in Prognostics and Health Management (PHM) by proposing and implementing a data-driven framework. The study posits that a system's characteristic life, represented by the Weibull scale parameter (λ), is not static but a dynamic variable influenced by time-varying operational and environmental factors, or covariates. Unlike traditional methods, such as the Accelerated Failure Time (AFT) model, which rely on restrictive linear assumptions to link covariates to the lifetime model, this methodology employs supervised Machine Learning (ML) models to quantify this complex and potentially non-linear relationship. The methodology is validated using the National Aeronautics and Space Administration (NASA) Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) aircraft engine degradation dataset. The results demonstrate that ensemble-based methods like Light Gradient Boosting Machine (LightGBM) and Extreme Gradient Boosting (XGBoost) significantly outperform classical statistical models across a suite of standard regression metrics and custom-developed Weibull-centric error metrics. The consistent performance rankings across all metrics validate the effectiveness of the proposed approach and confirm its capability to capture the intricate degradation dynamics of real-world systems.

Keywords-Prognostics and Health Management (PHM); weibull distribution; Remaining Useful Life (RUL); Accelerated Failure Time (AFT); machine learning; supervised learning

I. INTRODUCTION

Prognostics and Health Management (PHM) is a set of methodologies used for predictive maintenance, a strategy based on an asset's condition and the predicted future evolution of that condition [1]. Central to PHM is the estimation of Remaining Useful Life (RUL), defined as the time remaining until a failure occurs. Accurate RUL prediction is important for optimizing maintenance policies, ensuring system safety, and reducing costs. This prediction is complicated by the dynamic nature of RUL as it changes over time, and its inherent uncertainty, requiring it to be modeled as a random variable.

To address the need for robust and probabilistic modelling, the two-parameter Weibull distribution is widely used for reliability engineering and survival analysis due to its ability to model diverse failure characteristics [2]. Its shape parameter (k) describes the nature of the failure rate, while its scale parameter (λ) represents the system's characteristic life. Traditionally, these parameters are treated as static, which is a significant limitation for systems operating under variable conditions.

This paper addresses this limitation by suggesting that a system's characteristic life is not fixed but is a dynamic variable influenced by time-varying operational and environmental factors, known as covariates. The central challenge is that classical survival models, such as the Accelerated Failure Time (AFT) model, often rely on restrictive linear assumptions to link covariates to the lifetime model [3]. These assumptions may not capture the complex, non-linear dynamics inherent in real-world machinery degradation.

To overcome this, we propose the use of supervised Machine Learning (ML) models to directly quantify the complex, potentially non-linear relationship between sensor-based covariates and the Weibull scale parameter (λ). By learning the direct function from data, the framework creates a more adaptive and accurate RUL prediction. Specifically, we directly model the Weibull scale parameter (λ) as a dynamic function of time-varying covariates, treating the RUL as an output of the probabilistic Weibull distribution rather than a direct output of the ML model [4, 5]. This methodology is a direct implementation of concepts proposed in the literature, which suggested exploring parametric models that link Weibull parameters to sensor data [6]. Furthermore, by embedding the ML prediction within an established stochastic framework, this hybrid approach offers a novel solution to the Uncertainty Quantification (UQ) challenge identified in black-box approximations [7].

II. EXPERIMENTAL SETUP AND METHODOLOGY

A. Dataset

The dataset utilized in this research is the Aircraft Engine Degradation Simulation Data created using a high-fidelity Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) model that simulates damage propagation in aircraft gas turbine engines [8]. The dataset is divided into multiple subsets with different operating conditions. This study uses the "FD001" subset, which contains 100 training trajectories and 100 test trajectories representing a single operating condition at sea level. Each data file (train_FD001 and test_FD001) contains multivariate time series data, where each row represents a single snapshot of a specific engine's operational parameters at a given point in its lifecycle. The columns in the dataset represent the following variables:

- Engine Index: Identifies a unique aircraft engine instance.
- Cycle: The time step or cycle number for each engine.
- Operational Settings 1, 2, and 3: Three operational settings that serve as control inputs for the simulation.
- Sensor Measurements: Twenty-one sensor readings, capturing the engine's real-time performance and degradation.

The degradation is modeled as a loss of flow and efficiency in the engine modules, with an exponential rate of change that continues until a failure criterion is met. The provided RUL_FD001 file contains the corresponding RUL values for each of the 100 test trajectories in test_FD001, which serves as ground truth for model evaluation.

B. Weibull Distribution

The mathematical formulations of the two-parameter Weibull distribution are depicted in the Probability Density Function (PDF) in (1), and the Cumulative Distribution Function (CDF) in (2):

$$f(t,\lambda,k) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} e^{-\left(\frac{t}{\lambda}\right)^{k}} \quad (1)$$

$$F(t,\lambda,k) = 1 - e^{-\left(\frac{t}{\lambda}\right)^{k}} \quad (2)$$

where PDF gives the relative likelihood of failure at a specific time t, while CDF calculates the total probability that a system will have failed by time t. The parameter k is the shape parameter, which governs the failure rate pattern, with a value of $k > 1$ indicating wear-out degradation common to most machinery. The scale parameter λ defines the system's characteristic life, representing the time by which approximately 63.2% of units are expected to fail. A larger λ stretches the distribution along the time axis and corresponds to a longer expected life.

C. Models for Estimating the Covariate-Dependent Scale Parameter

This research assumes that a system's characteristic life, represented by the Weibull scale parameter λ, is not static but is a dynamic function of its time-varying covariates X, expressed as $\lambda = \lambda(X)$. For the aircraft engines case, covariates include three operational settings and twenty-one sensor measurements. To find the best approximation of this function, this study implements and compares two distinct modeling philosophies: a classical statistical model to serve as a baseline and a suite of supervised ML models.

A classical AFT model is first used as a baseline. The AFT model assumes that covariates accelerate or decelerate the failure time through a multiplicative factor and relies on a log-linear relationship between covariates and lifetime. Regression coefficients β quantifying this relationship are determined using Ordinary Least Squares (OLS) regression. Once trained, the model predicts the covariate-adjusted scale parameter by applying this acceleration factor to a baseline scale parameter λ_{base} using (3):

$$\lambda(X) = \lambda_{base} e^{-(\beta^T X)} \quad (3)$$

To address the limitations of rigid relationships of classical models, a supervised learning framework is used to learn the non-linear mapping from covariates to λ directly from data. Several algorithms are tested, including Linear Regression (LR) and Support Vector Regression (SVR). SVR is included due to its strengths in handling non-linear and high-dimensional relationships through kernel functions [10, 11].

Ensemble methods are also employed because they enhance predictive performance and improve generalization [12]. This study implements models from both main ensemble families. From the bagging family, Random Forest (RF) is selected as a popular, high-accuracy model that mitigates overfitting by averaging trees trained on random data subsets [13]. From the boosting family, Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM) are used, due to their regularized learning objective to avoid over-fitting and high accuracy prediction [14, 15], respectively.

A key element of this methodology is the design of the prediction target. Directly using the final RUL can introduce overfitting and distort the Weibull distribution. Instead, models are trained to predict a transformed target y, which represents a linear degradation function from the characteristic life λ to zero at the actual time of failure t_{RUL}:

$$y = \lambda\left(1 - \frac{t}{t_{RUL}}\right) \quad (4)$$

D. Data Preprocessing

1) Denoising

The sensor measurements in the raw dataset contain inherent noise, which can obscure the underlying degradation trend. To mitigate this, a smoothing algorithm based on the Simple Moving Average (SMA) method was applied to the sensor data, which smooths out random fluctuations in data by calculating the average value over a specified period using:

$$y_t = \frac{1}{w}\sum_{i=0}^{w-1} x_{t-i} \quad (5)$$

where y_t represents the smoothed value at time t, x is the original sensor data, and w is the window size. The rationale for using this method is that it effectively reduces short-term fluctuations and highlights longer-term trends, thereby making the engine's degradation signal clearer and more suitable for modeling.

2) Scaling

Scaling adjusts feature values to a common range and is known to improve model performance only if, as explained in [16], the technique employed is achieves appropriate scaling for the data used. For this study, the Standard Scaler was used, given in (6):

$$z = \frac{x-u}{s} \quad (6)$$

where, z is the standardized value, x is the original feature value, u is the mean of the training samples, and s is the standard deviation of the training samples. After scaling, all features are centered around zero with unit variance, preventing features with larger magnitudes from disproportionately influencing the training process.

E. Performance Metrics

1) Regression Metrics

To evaluate model performance, three common metrics are employed: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). MAE measures the average magnitude of the errors e_i in a set of predictions without considering their direction, as shown in:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |e_i| \quad (7)$$

MSE is the average of the squared differences between predicted and actual values, defined in (8). Squaring the errors gives greater penalties to larger errors, making MSE highly sensitive to outliers.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \quad (8)$$

Finally, the RMSE is the square root of the MSE, as shown in (9). RMSE has the advantage of being in the same unit as the target variable, making it more interpretable than MSE, while still penalizing large errors heavily.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \quad (9)$$

As noted in [17], there is an ongoing debate regarding the relative merits of MAE versus RMSE. Neither metric is inherently better, as their optimality depends on the error distribution: RMSE is optimal for normal (Gaussian) errors, while MAE is optimal for Laplacian errors. MAE is also widely considered a more robust form of evaluation, as it is less sensitive to extreme outliers than the squared-error terms in MSE and RMSE. By presenting both metrics, this study provides a multi-faceted view of the error, capturing both the average error magnitude using MSE and the impact of larger, undesirable errors using RMSE.

2) Weibull-Based Error Metrics

a) Weibull Error

Standard regression metrics such as MAE and RMSE are insufficient for evaluating probabilistic forecasts. While authors in [4] proposed metrics like Continuous Ranked Probability Score (CRPS) to evaluate the accuracy of predicted distribution, our work introduces a novel suite of event-based metrics, Mean Weibull Error (MWE) and Mean Quadratic Weibull Error (MQWE), designed to evaluate how well the entire predicted failure curve (Figure 1) matches the actual, discrete failure event, providing a more holistic assessment of prognostic performance.

The underlying concept is to measure the discrepancy between the continuous Weibull CDF, $F(t)$, and the actual failure outcome, which is represented as a step function that transitions from 0 to 1 at the failure time. The error corresponds to the area between these two curves, illustrated as the error area in Figure 1. The Weibull Error is defined in (10):

$$\mathrm{WE}(t,\lambda,k) = e_l + e_r \tag{10}$$

where

$$e_l = \int_0^t F(t)\,dt,\; e_r = \int_t^\infty 1 - F(t)\,dt$$

where t denotes the observed failure time, λ denotes the scale parameter k denotes the shape parameter, e_l calculates the area under the Weibull CDF from time 0 up to the actual failure time and penalizes non-zero predicted failure probability before the actual failure, and e_r calculates the area between the Weibull CDF and the value 1 from the actual failure time to infinity and penalizes failure probabilities below 1 after failure has occurred. An ideal model would yield zero for both terms. An analytical solution to this integral is given in (11):

$$\mathrm{WE} = t + \frac{2\lambda}{k}\Gamma\left(\frac{1}{k},\left(\frac{t}{\lambda}\right)^k\right) - \frac{\lambda}{k}\Gamma\left(\frac{1}{k}\right) \tag{11}$$

where the upper incomplete gamma function is used, denoted by $\Gamma(s,z)$. The Weibull Error for a dataset is computed as the mean of individual errors, as shown in (12):

$$\mathrm{MWE} = \frac{1}{n}\sum_{i=1}^{n} WE(t_i,\lambda,k) \tag{12}$$

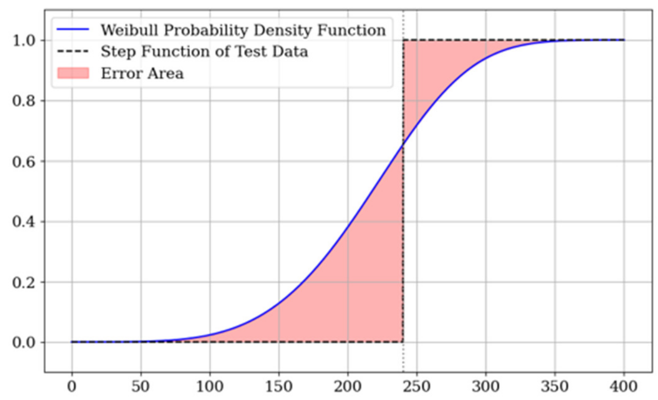

Fig. 1. Graph of error area.

b) Left Weibull Error

The Left Weibull Error (LWE) is a metric designed to quantify the model's performance in the period leading to an actual system failure. This error term specifically measures the cumulative probability of failure that the model predicts before the event occurs. It is represented by the area under the Weibull CDF from time 0 to the actual failure time t:

$$\mathrm{LWE}(t,\lambda,k) = e_l \tag{13}$$

Simplified to:

$$\mathrm{LWE} = t + \frac{\lambda}{k}\Gamma\left(\frac{1}{k},\left(\frac{t}{\lambda}\right)^k\right) - \frac{\lambda}{k}\Gamma\left(\frac{1}{k}\right) \tag{14}$$

A high LWE indicates that the model is prematurely forecasting a significant probability of failure, which can be interpreted as a penalty for false-positive failure predictions. An ideal model, which perfectly predicts the time of failure, would maintain a CDF value of 0 until the exact moment of failure, resulting in an LWE of zero. The dataset-level metric is computed as the mean of individual errors using (15):

$$\mathrm{MLWE} = \frac{1}{n}\sum_{i=1}^{n} WE(t_i,\lambda,k) \tag{15}$$

c) Right Weibull Error

In contrast, the Right Weibull Error (RWE) is a metric that measures the discrepancy between the model's predicted failure probability and the actual outcome, where the probability should be 1. This error term penalizes the model for being "too optimistic" or hesitant in its prediction.

The RWE is defined as the area between the Weibull CDF and the value of 1, starting from the point of actual system failure, t, and extending to infinity:

$$\mathrm{RWE}(t,\lambda,k) = e_l \tag{16}$$

Simplified to:

$$\mathrm{RWE} = \frac{\lambda}{k}\Gamma\left(\frac{1}{k},\left(\frac{t}{\lambda}\right)^k\right) \tag{17}$$

An ideal model would have its CDF jump to 1 immediately at time t, making this error term equal to zero. The Mean Right Weibull Error (MRWE) is calculated as the mean of individual errors:

$$\mathrm{MRWE} = \frac{1}{n}\sum_{i=1}^{n} RWE(t_i,\lambda,k) \tag{18}$$

d) Quadratic Weibull Error

The Squared Weibull Error increases the penalty for large deviations by squaring the discrepancy between the predicted CDF and the true failure event. This behavior is analogous to MSE and makes the metric sensitive to substantial prediction errors. The integral formulation is given in (19):

$$\mathrm{QWE} = \int_0^t (F(t))^2 dt + \int_t^\infty (1 - F(t))^2 dt \tag{19}$$

where the first term penalizes high predicted failure probabilities before failure, while the second penalizes slow convergence to a probability of 1 after failure. A closed-form analytical solution is provided in (20):

$$\mathrm{QWE} = t + \frac{2\lambda}{k}\Gamma\left(\frac{1}{k},\left(\frac{t}{\lambda}\right)^k\right) - \frac{\lambda}{k}\Gamma\left(\frac{1}{k}\right) + \frac{\lambda}{k2^{1/k}}\Gamma\left(\frac{1}{k}\right) \tag{20}$$

The dataset-level Mean Quadratic Weibull Error (MQWE) is computed as shown in (21):

$$\mathrm{MQWE} = \frac{1}{n}\sum_{i=1}^{n} QWE(t_i,\lambda,k) \tag{21}$$

F. Methodology

A structured workflow was followed, as illustrated in Figure 2, comprising parameter estimation, data preprocessing, model training, hyperparameter optimization, prediction, and evaluation.

The initial shape k and scale λ parameters were estimated using Maximum Likelihood Estimation (MLE) [18]. The shape parameter k is obtained by solving the implicit equation shown in (22):

$$\frac{1}{k} = \frac{\sum_{i=1}^{n} t_i^k \ln(t_i)}{\sum_{i=1}^{n} t_i^k} + \frac{1}{n}\sum_{i=1}^{n} \ln(t_i) \quad (22)$$

This equation does not admit a closed-form solution and was therefore solved using an iterative numerical procedure. The scale parameter was computed using:

$$\lambda = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^k\right)^{\frac{1}{k}} \quad (23)$$

For the dataset used in this work, the estimated parameters are k = 4.359886822599401 and λ = 236.7659325811420. The value of $k > 1$ signifies a consistent wear-out phase in the aircraft engine's degradation, with a characteristic life centered around λ cycles.

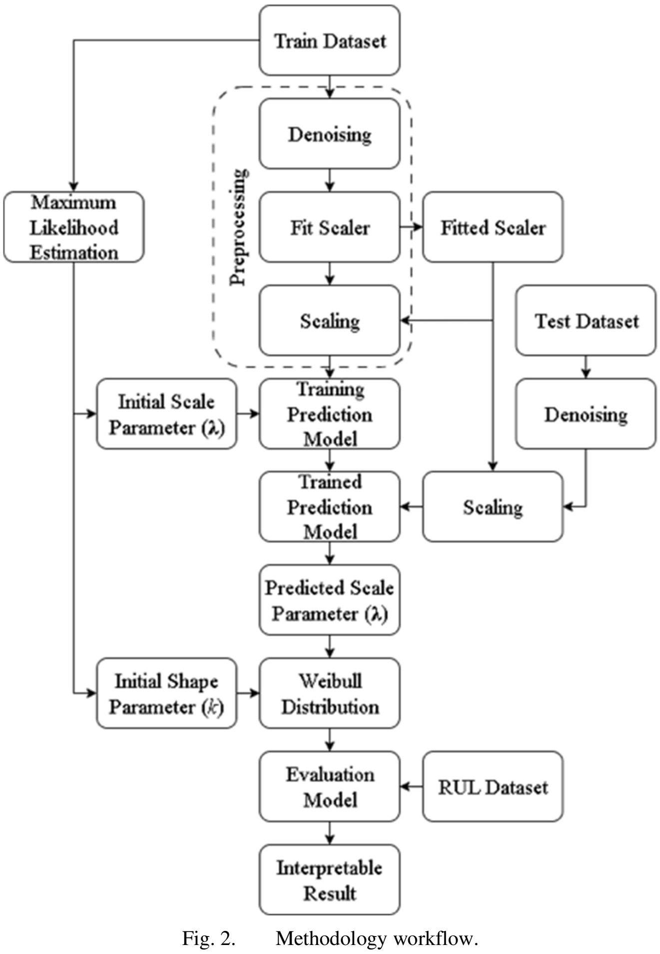

Fig. 2. Methodology workflow.

Then, both training and test datasets underwent an identical preprocessing pipeline. Following preprocessing, the modeling phase was initiated. Each prediction model was trained to estimate a covariate-dependent Weibull scale parameter using the processed feature matrix. The estimated scale parameter was then combined with the fixed shape parameter (obtained via MLE) to construct a complete Weibull distribution for each test instance.

Since ML model performance is highly sensitive to hyperparameter selection, to ensure optimal and reproducible results, hyperparameter tuning was conducted using a grid search strategy [19, 20]. This approach systematically evaluates all possible combinations within predefined parameter grids to identify the configuration that minimizes validation error. The AFT model was included as a classical statistical baseline and followed a parallel prediction workflow. The fitted model was then used to compute a covariate-adjusted scale parameter $\lambda(X)$, enabling a direct and fair comparison with the data-driven approaches.

In the final evaluation phase, predicted Weibull distributions were compared against ground-truth failure information derived from the RUL dataset. Additionally, for each test engine, the true failure time was reconstructed by adding the last observed operational cycle to the corresponding ground-truth RUL value provided in the RUL_FD001 file. This reconstruction ensures that evaluation is performed over a complete lifecycle, incorporating both the observed degradation history and the actual failure point.

III. RESULTS AND DISCUSSION

The performance standard regression metrics and the proposed Weibull-based error metrics of the models employed for the prediction of the Weibull scale parameter λ based on system covariates are presented in Table I. The No Covariate (NC) baseline model serves as a reference point, representing a traditional, static Weibull model. Its performance highlights the significant limitations of a static approach when dealing with systems operating under variable conditions. Additionally, the fact that all other models, particularly those based on ML, consistently achieved lower error metrics i) demonstrates the value of incorporating covariates, and ii) confirms the hypothesis that a system's characteristic life is not fixed but is a dynamic variable influenced by time-varying factors.

The ensemble-based ML models, namely LightGBM, XGBoost, and RF, exhibited superior performance across all evaluated metrics. LightGBM emerged as the top-performing model, achieving the lowest errors in all standard regression metrics (MAE=25.44 cycles, MSE=1170.52 cycles^2, RMSE=34.21 cycles) and the Weibull-based metrics (MWE=27.16 cycles, MQWE=1058.44 cycles). In addition, the low MQWE achieved by LightGBM indicates that its predicted Weibull CDF curve most closely matches the ideal step function of the actual failure event, penalizing large errors and ensuring a precise fit. The strong performance of these models confirms the central hypothesis of this research: that supervised ML models are highly effective at capturing the complex, non-linear relationships between covariates and the Weibull scale parameter, providing a more robust and accurate RUL prediction. In contrast, the classical statistical models yielded less favorable results. The AFT model with OLS regression (AFTOLS) performed worse than all ML models and only slightly better than the NC baseline model. This suggests that the restrictive linear assumptions of the AFT model are inadequate for capturing the intricate degradation dynamics of the aircraft engine, highlighting the limitations of traditional methods in this context. The LR and SVR models, while outperforming the classical statistical baselines, were consistently surpassed by the ensemble methods.

TABLE I. THE RESULT OF PERFORMANCE COMPARISON BETWEEN METHODS

	MAE (cycles)	**MSE (cycles^2)**	**RMSE (cycles)**	**MWE (cycles)**	**MLWE (cycles)**	**MRWE (cycles)**	**MQWE (cycles)**
NC	38.50	2553.77	50.53	43.18	10.83	32.35	2279.94
AFTOLS	36.17	2366.02	48.64	44.09	10.68	33.40	2427.67
RF	26.62	1212.89	34.83	28.20	9.78	18.43	1107.17
XGBoost	26.59	1230.37	35.08	27.86	9.03	18.83	1100.96
LightGBM	25.44	1170.52	34.21	27.16	9.07	18.09	1058.44
SVR	29.74	1357.97	36.85	30.56	11.90	18.67	1256.38
LR	35.43	1706.57	41.31	34.37	9.56	24.82	1448.33

Moreover, the developed Weibull-based error metrics, including MWE, Mean Left Weibull Error (MLWE), MRWE, and MQWE, showed a strong correlation with the results from standard regression metrics. The models that ranked highest on MAE and RMSE also ranked highest on the custom metrics. This consistency serves as a validation of the custom metrics, confirming their effectiveness as reliable tools for assessing model performance in Weibull-based survival analysis.

IV. CONCLUSION

This study introduced a new hybrid framework that combines supervised Machine Learning (ML) with classical Weibull survival analysis to address a significant gap in Prognostics and Health Management (PHM). The framework directly links Weibull model parameters to sensor data, offering a structured solution to Uncertainty Quantification (UQ) challenges by embedding ML predictions within an interpretable probabilistic Weibull framework, rather than relying on black-box methods like Monte Carlo dropout. The study compared this data-driven approach to traditional statistical models and validated its performance using both standard and custom Weibull-centric error metrics, such as Mean Weibull Error (MWE) and Mean Quadratic Weibull Error (MQWE), that evaluate how well the predicted failure curves match discrete failure events, which are crucial for maintenance decisions.

The results of the experimental comparison demonstrated that ML models, particularly ensemble-based methods like Light Gradient Boosting Machine (LightGBM) and Extreme Gradient Boosting (XGBoost), significantly outperformed classical statistical models, including the Accelerated Failure Time model with Ordinary Least Squares regression (AFTOLS). This finding validates the hypothesis that traditional models, which rely on rigid linear assumptions, are often inadequate for capturing the complex, non-linear degradation dynamics of real-world systems. Moreover, the performance of the ML models shows their capability to learn a more accurate and robust relationship between system covariates and the dynamic Weibull scale parameter.

In conclusion, this study addresses the limitations of conventional, static survival models and offers a powerful alternative for PHM applications, leading to a more adaptive and accurate Remaining Useful Life (RUL) prediction.

Future work could explore the application of this methodology to other datasets and the integration of more advanced ML techniques to further enhance predictive accuracy and robustness. For future work, it would be beneficial to explore alternative estimation methods for the AFT model, such as Maximum Likelihood Estimation (MLE), to determine if they can yield a more precise fit than the OLS method used in this study. Additionally, incorporating other established classical survival models, such as the Proportional Hazards model, would provide a more comprehensive baseline for comparison against the ML approach.

REFERENCES

[1] P. Dersin, *Modeling Remaining Useful Life Dynamics in Reliability Engineering*, 1st ed. Boca Raton: CRC Press, 2023.

[2] N. L. Clement and R. C. Lasky, "Weibull Distribution and Analysis: 2019," in *2020 Pan Pacific Microelectronics Symposium (Pan Pacific)*, HI, USA, Feb. 2020, pp. 1–5, https://doi.org/10.23919/PanPacific48324.2020.9059313.

[3] R. Saikia and M. P. Barman, "A Review on Accelerated Failure Time Models," *International Journal of Statistics and Systems*, vol. 12, no. 2, pp. 311–322, 2017.

[4] I. De Pater and M. Mitici, "Novel Metrics to Evaluate Probabilistic Remaining Useful Life Prognostics with Applications to Turbofan Engines," *PHM Society European Conference*, vol. 7, no. 1, pp. 96–109, June 2022, https://doi.org/10.36001/phme.2022.v7i1.3320.

[5] M. Mitici, I. De Pater, A. Barros, and Z. Zeng, "Dynamic predictive maintenance for multiple components using data-driven probabilistic RUL prognostics: The case of turbofan engines," *Reliability Engineering & System Safety*, vol. 234, June 2023, Art. no. 109199, https://doi.org/10.1016/j.ress.2023.109199.

[6] M. Salinas-Camus and N. Eleftheroglou, "Uncertainty in Aircraft Turbofan Engine Prognostics on the C-MAPSS Dataset," *PHM Society European Conference*, vol. 8, no. 1, June 2024, Art. no. 10, https://doi.org/10.36001/phme.2024.v8i1.4007.

[7] M. Dhada, L. Bull, M. Girolami, and A. Parlikad, "Collaborative prognosis using a Weibull statistical hierarchical model," *Reliability Engineering & System Safety*, vol. 262, Oct. 2025, Art. no. 111110, https://doi.org/10.1016/j.ress.2025.111110.

[8] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *2008 International Conference on Prognostics and Health Management*, Denver, CO, USA, Oct. 2008, pp. 1–9, https://doi.org/10.1109/PHM.2008.4711414.

[9] D. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 140–147, Dec. 2020, https://doi.org/10.38094/jastt1457.

[10] Z. Jun, "The Development and Application of Support Vector Machine," *Journal of Physics: Conference Series*, vol. 1748, no. 5, Jan. 2021, Art. no. 052006, https://doi.org/10.1088/1742-6596/1748/5/052006.

[11] A. F. Jadama and M. K Toray, "Ensemble Learning: Methods, Techniques, Application," Preprint, 2024, https://doi.org/10.13140/RG.2.2.28017.08802.

[12] Z. Zhu, "Systematic Optimization of Overfitting Problem in Machine Learning," *Highlights in Science, Engineering and Technology*, vol. 111, pp. 353–359, Aug. 2024, https://doi.org/10.54097/3tkzrj84.

[13] H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, June 2024, https://doi.org/10.58496/BJML/2024/007.

[14] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA, Aug. 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[15] V. S. F. Gonçalves and V. R. De Carvalho, "A Review of Interpretability Methods for Gradient Boosting Decision Trees," *Journal of the Brazilian Computer Society*, vol. 31, no. 1, pp. 640–654, Aug. 2025, https://doi.org/10.5753/jbcs.2025.5324.

[16] L. B. V. De Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performance," *Applied Soft Computing*, vol. 133, Jan. 2023, Art. no. 109924, https://doi.org/10.1016/j.asoc.2022.109924.

[17] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geoscientific Model Development*, vol. 15, no. 14, pp. 5481–5487, July 2022, https://doi.org/10.5194/gmd-15-5481-2022.

[18] N. A. M. Ikbal, S. A. Halim, and N. Ali, "Estimating Weibull Parameters Using Maximum Likelihood Estimation and Ordinary Least Squares: Simulation Study and Application on Meteorological Data," *Mathematics and Statistics*, vol. 10, no. 2, pp. 269–292, Mar. 2022, https://doi.org/10.13189/ms.2022.100201.

[19] M. Mwita, J. Mbelwa, J. Agbinya, and A. Elikana Sam, "The Effect of Hyperparameter Optimization on the Estimation of Performance Metrics in Network Traffic Prediction using the Gradient Boosting Machine Model," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10714–10720, June 2023, https://doi.org/10.48084/etasr.5548.

[20] T. Yu and H. Zhu, "Hyper-Parameter Optimization: A Review of Algorithms and Applications." arXiv, Mar. 2020, https://doi.org/10.48550/arXiv.2003.05689.