

Interpretable Failure Rate Prediction for Grouped Smart Meters using Bayesian Network-enhanced Survival Modeling

Zhengbo Zhang, Chuanji Zhang, Cheng He, *Senior, IEEE*, Yuntian Zhou, Tianfu Huang, and Hongbin Li

Abstract—The large-scale deployment of smart meters in power grids necessitates accurate prediction of failure rates to enable preventive maintenance. However, existing prediction models are often limited by insufficient multi-factor modeling and weak interpretability. To address these limitations, we propose a Bayesian network (BN)-enhanced Cox proportional hazards model (CPHM) for failure prediction of grouped smart meters. Multiple source covariates, including operational lifetimes, verification records, production batches, deployment areas, and component parameters, are jointly incorporated. Smart meters are first grouped by production batch and deployment area to capture intra-group homogeneity and inter-group heterogeneity in failure behavior. A global BN is then trained on the entire dataset to model probabilistic dependencies among covariates and quantify their risk contributions, thereby enhancing interpretability. Specifically, the conventional Cox linear predictor is replaced by a BN-enhanced risk function that incorporates conditional dependencies among covariates. Based on the learned BN, group-specific CPHMs with nonparametric baseline hazard functions are constructed to characterize distinct failure dynamics across groups. Validation on operational data from 38,238 smart meters, spanning 24 production batches and 358 deployment areas, shows that the proposed method achieves a C-index of 0.95, an R^2 of 0.90, and a mean time-dependent AUC of 0.97, outperforming state-of-the-art survival models. The results demonstrate strong capability in handling censored data and heterogeneous failure characteristics.

Index Terms—Smart metering, Cox proportional hazards model, failure rate prediction.

NOMENCLATURE

$\mathbf{X} = \{X_1, \dots, X_p\}$	Set of p random covariates
$\mathbf{x} = (x_1, \dots, x_p)^T$	Observed covariate vector
$\mathbf{x}^{(i)}$	Covariate vector of the i -th sample
$\text{Pa}(X_i)$	Parent set of node X_i in the BN
$P(X_i \text{Pa}(X_i))$	Conditional probability of X_i
T	Survival time
$h(t \mathbf{x})$	Hazard function (instantaneous failure rate)
$h_0(t)$	Baseline hazard function
$h_{0,g}(t)$	Baseline hazard of group g

This work was supported by the National Key Research and Development Program of China (No. 2022YFB2403800), and the Key Research and Development Program of Hubei Province (No. 2024BAB091). The computation is completed on the HPC Platform of Huazhong University of Science and Technology.

Cheng He is the corresponding author.

Z. Zhang, C. He, C. Zhang, Y. Zhou, and H. Li are with the State Key Laboratory of Advanced Electromagnetic Technology, School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074 China (e-mail: chenghe_seee@hust.edu.cn).

T. Huang is with the Marketing Service Center of State Grid Fujian Electric Power Co., Ltd.

$H(t \mathbf{x})$	Cumulative hazard function
$S(t \mathbf{x})$	Survival function
$F(t)$	Cumulative failure probability
$\beta = (\beta_1, \dots, \beta_p)^T$	Risk-weight vector
$f(\mathbf{x}; \beta)$	BN-enhanced risk function
$L(\beta)$	Partial likelihood
$\mathcal{L}(\beta)$	Weighted negative log-partial likelihood
d_τ	Number of failure events at time τ
$R(\tau)$	Risk set at time τ
$D(\tau)$	Set of failure events at time τ
K	Total number of groups
N_g	Number of samples in group g
$\mathcal{J}_{\text{inter}}, \mathcal{J}_{\text{intra}}$	Index sets of inter-/intra-group covariates
J	Number of inter-group covariates
$f_{\text{inter}}, f_{\text{intra}}$	Inter-/intra-group risk function components
$\mathcal{E}(X_j)$	Normalized explanatory power of covariate X_j

I. INTRODUCTION

Smart meters are intelligent electricity measurement devices that serve as the core sensing and data-acquisition layer of smart grids. They generally comprise several units, i.e., sensor, sampling, control, and communication, and continuously record the real-time information, including voltage, current, frequency, and electricity consumption [1]. Meanwhile, advances in communication networks and data management platforms have enabled bidirectional information exchange between the smart grid operators and users [2]. As a result, smart meters play an essential role not only in ensuring the accuracy, fairness, and transparency of power transactions, but also in supporting data-driven applications, e.g., demand-response potential estimation, dynamic price design, and consumer behavior profiling [3]. The rapid development of smart grids inspires the deployment of smart meters worldwide. The global installed base has exceeded 1.2 billion units, and the cumulative number of installed smart meters in China had surpassed 660 million by the end of 2024 [4].

As smart meters continue to integrate more complex electronic components and embedded functions, their system complexity increases, and their reliability requirements become more stringent [5]. More specifically, their operational reliability is affected by multiple factors, including component aging, environmental conditions, manufacturing inconsistencies, and diverse operating scenarios. In large-scale deployments, smart meter failures remain a practical concern for both utilities and users, including battery degradation, clock malfunctions,

and measurement inaccuracies in regional power systems [6]. Dynamic Bayesian networks (BNs) have also been applied to model the degradation processes of smart meter functional modules for reliability prediction [7].

Existing studies mainly rely on routine inspection and post-fault maintenance, which are often insufficient for large-scale smart meter systems. By contrast, failure rate prediction directly reflects the degradation and health conditions of smart meters [8]. It enables targeted preventive maintenance before widespread failures occur, thereby improving the long-term reliability of smart metering systems [9]. To address reliability concerns, many data-driven methods have been developed for failure assessment [10]. Simultaneously, trustworthy data-driven frameworks that explicitly incorporate domain knowledge into learning have been shown to improve both prediction accuracy and reliability in power system applications [11]. Earlier studies mainly relied on statistical methods and machine learning models, whereas more recent work has introduced deep learning techniques to improve prediction accuracy [12]. For example, the Transformer-Encoder-BiLSTM model achieves high accuracy in operating-state classification, but its black-box nature limits interpretability and reduces its practicality [13]. A similar challenge exists in power system stability assessment, where deep learning models achieve strong predictive accuracy but often lack the physical reliability required for operational decision-making [14]. In addition, information-fusion methods based on entropy weighting can improve robustness in multivariate analysis, though they generally require extensive data preprocessing and carefully designed input features [15].

Survival analysis has been widely used in smart meter failure modeling, attributed to its explicit description of the time-to-failure process. In particular, the Cox proportional hazards model (CPHM) provides clear interpretations of how different factors influence the risk of failure and can naturally handle censored data (incomplete failure information due to some assets not failing within the study period) [16]. Neural network-based survival models have recently been developed to handle censored degradation data through novel loss functions derived from survival likelihoods [17]. Recent work has further improved the CPHM by incorporating machine-learning-derived covariates to better capture equipment degradation under varying operating conditions [18]. Despite its benefits, CPHM-based methods generally rely on a single regression structure and do not explicitly account for heterogeneous subgroup effects. This limitation reduces their ability to describe failure variations associated with production batches, deployment environments, and operating conditions [19]. Other techniques, such as gradient-boosting decision trees [20], online extended Kalman filtering [21], and AGG-ARIMA models [22], have also been applied to fault identification, real-time estimation, and small-sample learning. Metaheuristic-optimized machine learning approaches have also shown promise in automated feature selection for fault detection in complex systems [23]. Nevertheless, these methods still show limitations in interpretability, robustness, or computational efficiency [24].

Another challenge is that smart meters are deployed in large, heterogeneous groups, whereas failure observations are sparse

due to their inherent reliability [25]. These characteristics make failure rate prediction more difficult and reduce the practical applicability of conventional methods. Therefore, there remains a need for a modeling framework that can simultaneously achieve high prediction accuracy, integrate multiple factors effectively, and provide clear interpretability for large-scale smart metering systems. To address this issue, we propose a BN-enhanced CPHM framework for predicting the failure rates of grouped smart meters. The main new contributions are summarized as follows:

- A BN-enhanced risk function is developed by incorporating BN-derived conditional probabilities into the Cox hazard model. The resulting dependency-adjusted representation enables the risk score to capture both covariate magnitudes and probabilistic dependencies, while preserving coefficient-level interpretability.
- A global–local modeling architecture is proposed, in which a shared BN and a set of risk-weight parameters are learned from the full dataset, while group-specific non-parametric baseline hazards are constructed to characterize heterogeneous failure dynamics. A weighted negative log-partial likelihood is further introduced to balance the influence of groups with different sample sizes, thereby improving generalization across heterogeneous operating conditions.
- An interpretable failure-risk analysis mechanism is developed to explain both group-level and individual-level failure behavior. At the group level, a posterior-based attribution method identifies the dominant covariates associated with baseline hazard peaks. At the individual level, the instantaneous hazard is decomposed into inter-group and intra-group components, enabling quantitative interpretation of covariate contributions to failure risk.

The remainder of this paper is organized as follows. Section II gives the preliminaries. The details of the proposed method are then elaborated in Section III. Experimental results and discussions are given in Section IV, and conclusions are drawn in Section V.

II. PRELIMINARIES

A. Bayesian Network

A BN is a probabilistic graphical model widely adopted for reliability assessment across various engineering domains [26], [27]. Let $\mathbf{X} = \{X_1, \dots, X_p\}$ denote the set of p covariates. In this study, uppercase bold $\mathbf{X}$ denotes the set of random variables (covariates), while lowercase bold $\mathbf{x}$ denotes the corresponding observed values. Each node corresponds to one covariate, and an edge $X_j \rightarrow X_i$ indicates that X_i directly depends on X_j . The parent set of X_i is

$$\text{Pa}(X_i) = \{X_j \mid X_j \rightarrow X_i\}. \quad (1)$$

Accordingly, the joint probability distribution of $\mathbf{X}$ is

$$P(\mathbf{X}) = \prod_{i=1}^p P(X_i \mid \text{Pa}(X_i)). \quad (2)$$

For an observed covariate vector $\mathbf{x} = (x_1, \dots, x_p)^T$, the BN provides conditional probabilities $P(X_j = x_j \mid \text{Pa}(X_j))$,

which quantify the consistency of each observed covariate with the learned dependency structure. In this work, the BN structure and its conditional probability tables (CPTs) are learned using hill-climbing with the Bayesian information criterion (BIC).

B. Cox Proportional Hazards Model

The CPHM is a survival analysis model for time-to-event data [28]. Let T denote the survival time, defined as the elapsed time from meter verification to removal. If a meter has not failed by the end of the observation period t , the corresponding observation is treated as right-censored. The hazard function is defined as

$$h(t | \mathbf{x}) = h_0(t) \exp(\beta^T \mathbf{x}), \quad (3)$$

where $h_0(t)$ is the baseline hazard function, $\mathbf{x} = (x_1, \dots, x_p)^T$ is the covariate vector, and $\beta = (\beta_1, \dots, \beta_p)^T$ is the regression coefficient vector. The sign of β_j indicates whether the corresponding covariate increases or decreases failure risk. The cumulative hazard and survival functions are

$$H(t | \mathbf{x}) = H_0(t) \exp(\beta^T \mathbf{x}), \quad (4)$$

$$S(t | \mathbf{x}) = \exp(-H(t | \mathbf{x})). \quad (5)$$

Model parameters are estimated by maximizing the partial likelihood

$$L(\beta) = \prod_{i \in D(\tau)} \frac{\exp(\beta^T \mathbf{x}^{(i)})}{\sum_{j \in R(\tau)} \exp(\beta^T \mathbf{x}^{(j)})}, \quad (6)$$

where $R(\tau)$ and $D(\tau)$ denote the risk set and the set of failure events at time τ , respectively. Since the partial likelihood depends only on β and the ordering of failure times, it enables consistent and asymptotically normal estimation of β without specifying $h_0(t)$. After β is estimated, the baseline cumulative hazard is obtained via the Breslow estimator [29] as

$$\hat{H}_0(t) = \sum_{i \in D(\tau), \tau \leq t} \frac{d_\tau}{\sum_{j \in R(\tau)} \exp(\beta^T \mathbf{x}^{(j)})}, \quad (7)$$

where d_τ denotes the number of failure events observed at time τ . The proportional hazards assumption can be checked using cumulative sums of martingale-based residuals [30].

C. Failure Rate and Failure Probability

Let $F(t) = P(T \leq t)$ denote the cumulative failure probability, with density function $f(t)$, i.e., $F(t) = \int_0^t f(\tau) d\tau$. The failure rate, or instantaneous failure intensity, is $\lambda(t) = f(t)/(1 - F(t))$. Within the CPHM framework, the failure rate is equivalent to the hazard function: $\lambda(t | \mathbf{x}) = h(t | \mathbf{x})$ [31], which represents the instantaneous failure risk of smart meters still operating at time t . Therefore, failure rate prediction is formulated as a hazard estimation problem.

III. PROPOSED METHOD

Details of the proposed BN-enhanced CPHM framework for smart meter failure rate prediction are shown in Fig. 1. It integrates multi-source covariates through a BN and models time-to-failure behavior using group-specific CPHMs, which mainly comprises data preparation and grouping, BN-enhanced risk modeling, base CPHMs, and a three-stage training process for model estimation. The framework is built upon the proportional hazards assumption inherent in the CPHM and the conditional independence structure encoded by the BN, and its estimation procedure involves explicit grouping, dependency-aware risk modeling, and sequential stage-wise parameter learning, each with clearly defined inputs and outputs.

To begin with, a structured dataset is constructed from the original smart meter records, and the meters are grouped to explicitly represent both manufacturing quality and operating environment. Then, a BN is learned from the full dataset to characterize dependency relationships among multiple influencing factors. The BN structure, its CPTs, and the risk-weight parameters β are learned jointly from the full dataset and shared across all groups as global components. Based on the learned BN, a risk score function is established to describe the combined effect of covariates on failure risk. Next, a base CPHM is constructed for each smart meter group to estimate its group-specific baseline hazard curve. Each group maintains its own baseline hazard function $h_{0,g}(t)$ as a local component, reflecting the fact that temporal failure patterns differ due to manufacturing and deployment heterogeneity. Finally, by combining the baseline hazard with the BN-enhanced risk score, the failure rate and failure probability of each smart meter can then be estimated. The resulting global-local modeling architecture achieves statistical robustness by pooling all samples for BN structure learning and β estimation, which would not be attainable from small within-group samples alone; meanwhile, the group-specific baselines preserve each group's distinct failure dynamics without being diluted by cross-group averaging.

A. Data Preparation and Grouping

To build a reliable prediction model, multi-source data are first integrated and preprocessed. Our previous work, based on component analysis and simulation, indicates that four categories are most relevant, i.e., basic meter attributes, verification records, group identifiers, and component specifications [32]. For each smart meter, survival time is defined as the interval from verification to removal, and right-censored observations are retained to preserve incomplete lifetime information.

The covariates include both categorical variables, such as component-related attributes, and continuous variables, such as verification records. Among the continuous covariates, those exhibiting severely right-skewed distributions—specifically, Inventory Time and Inventory Time Ratio—are log-transformed prior to model fitting to improve normality and stabilize the risk estimation. For BN parameter estimation, the continuous covariates are first discretized. The categorical covariates are then encoded as one-hot vectors to support subsequent risk calculations.

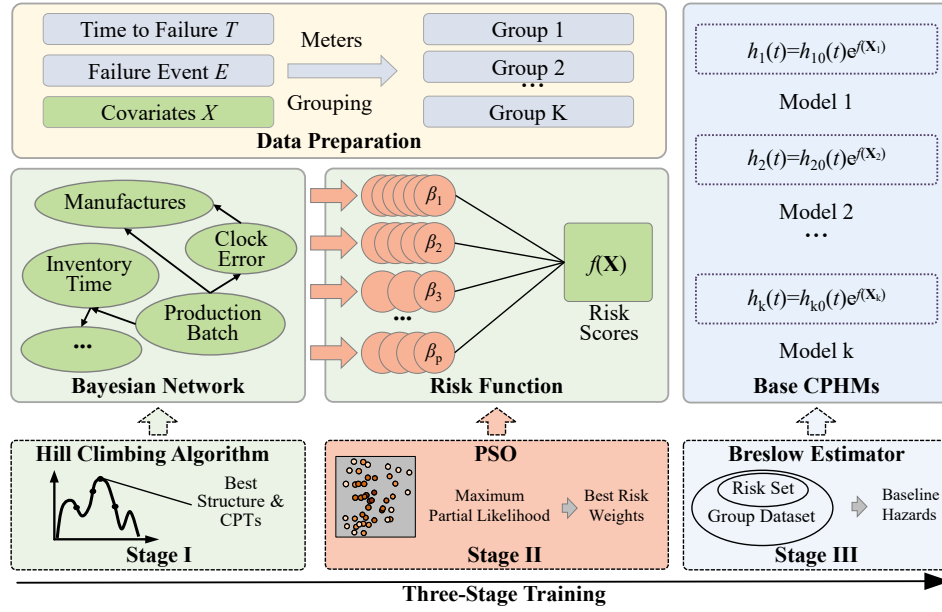


Fig. 1: Flowchart of the proposed BN-enhanced CPHM framework. The framework consists of data preparation and grouping, Bayesian network construction, and base CPHMs, and is trained through a three-stage process.

After preprocessing, smart meters are grouped according to production batch and deployment area. Meters within the same group are assumed to share similar manufacturing characteristics and operating environments, and therefore to exhibit a common baseline hazard function, enabling the use of group-specific CPHMs with sufficient within-group sample sizes for stable baseline estimation. To ensure sufficient statistical support for within-group estimation, a minimum group size of $N_g \geq 40$ is imposed.

Finally, the final dataset contains 38,238 single-phase smart meters from 24 production batches and 358 deployment areas. This grouping strategy enables heterogeneous failure dynamics to be captured through group-specific baseline hazard functions within the proposed BN-enhanced CPHM framework.

B. BN-enhanced Risk Function

The proposed BN-enhanced CPHM employs a dependency-aware risk function to characterize the effect of covariates on failure risk. Instead of using a simple linear predictor, the risk function incorporates the conditional dependency structure learned by the BN, thereby accounting for both covariate values and their probabilistic relationships in risk estimation.

The BN-enhanced risk function is defined as

$$f(\mathbf{x}; \boldsymbol{\beta}) = \sum_{j=1}^p \beta_j x_j P(X_j = x_j | \text{Pa}(X_j)), \quad (8)$$

where $\mathbf{x} = (x_1, \dots, x_p)^T$ denotes the observed covariate vector and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ denotes the corresponding risk-weight parameters. Here, the product $\beta_j x_j$ in each summand is understood as an inner product $\boldsymbol{\beta}_j^T \mathbf{x}_j$ when the j -th covariate is categorical: $\mathbf{x}_j \in \{0, 1\}^{d_j}$ is its one-hot encoding and $\boldsymbol{\beta}_j \in \mathbb{R}^{d_j}$ is a risk-weight vector of the same dimension, so $\boldsymbol{\beta}_j^T \mathbf{x}_j$ selects exactly one component of $\boldsymbol{\beta}_j$, yielding a scalar. When the j -th

covariate is continuous, the formulation degenerates to scalar product $\beta_j x_j$ with $d_j = 1$. Consequently, the contribution of each covariate is adjusted by its conditional probability under the BN, thereby enabling risk estimation that reflects both feature magnitude and dependency structure.

The term $x_j P(X_j = x_j | \text{Pa}(X_j))$ can be regarded as a dependency-adjusted covariate, where the original covariate value is weighted by its conditional probability under the learned BN structure. In this way, the risk function reflects not only the magnitude of each covariate but also its dependence on other related covariates. Thus, we provide a more informative description of multi-factor failure mechanisms while preserving interpretability at the coefficient level.

The term β_j represents the marginal contribution of the j -th dependency-adjusted covariate to failure risk. A positive β_j indicates that the corresponding covariate increases failure risk, whereas a negative β_j indicates a risk-reducing effect. Therefore, the estimated coefficients can still be directly interpreted in terms of covariate influence.

Moreover, as the covariate dependency structure does not vary across groups, the framework uses a single BN learned from the entire training set, along with a single set of risk-weight parameters $\boldsymbol{\beta}$ shared by all groups. To capture heterogeneous failure behavior, base CPHMs are constructed for each smart meter group, whose hazard function is

$$h_g(t | \mathbf{x}) = h_{0,g}(t) \exp(f(\mathbf{x}; \boldsymbol{\beta})), \quad (9)$$

where $h_{0,g}(t)$ is the group-specific baseline hazard function. Thus, the BN-enhanced risk function captures the individual effect of covariates, whereas the group-specific baseline hazard characterizes the shared failure tendency within each group.

With a fixed BN structure, the dependency-adjusted covariates are deterministic transformations of the observed inputs, so identifiability is retained under the standard regularity

conditions of the CPHM. Meanwhile, the exponential link ensures that the hazard remains monotonic with respect to $f(\mathbf{x}; \beta)$, thereby preserving the proportional hazards property. Though the proposed formulation is more expressive than the conventional CPHM, its complexity is effectively constrained by BIC-based BN structure learning, an unchanged number of risk-weight parameters, and group-wise partial-likelihood estimation with sample-size normalization.

C. Base CPHMs

Based on the BN-enhanced risk function, a base CPHM is constructed for each group of smart meters. For a smart meter with covariate vector $\mathbf{x}$ in group g , the cumulative hazard function is

$$\hat{H}_g(t | \mathbf{x}) = \hat{H}_{0,g}(t) \exp(f(\mathbf{x}; \hat{\beta})). \quad (10)$$

The corresponding instantaneous hazard, which is equivalent to the failure rate, is

$$\hat{\lambda}_g(t | \mathbf{x}) = \hat{h}_{0,g}(t) \exp(f(\mathbf{x}; \hat{\beta})). \quad (11)$$

Accordingly, the survival probability and failure probability can be obtained as

$$\hat{S}_g(t | \mathbf{x}) = \exp(-\hat{H}_g(t | \mathbf{x})), \quad (12)$$

$$\hat{F}_g(t | \mathbf{x}) = 1 - \hat{S}_g(t | \mathbf{x}). \quad (13)$$

Then, the group-specific baseline hazard captures the common failure tendency of smart meters within each group, whereas the BN-enhanced risk function captures the individual effects of covariates. Their combination enables individualized estimation of failure rate, survival probability, and failure probability.

D. Model Training

The proposed BN-enhanced CPHM framework is trained in three stages, as illustrated in Fig. 1. The three stages learn the BN structure, estimate the shared risk-weight parameters, and sequentially derive the group-specific baseline hazards, respectively.

In Stage I, the BN structure is learned from the training data using a score-based hill-climbing algorithm with the BIC as the scoring criterion. Once the network structure is determined, the corresponding CPTs are estimated. The search space consists of all directed acyclic graphs over the p covariates, and the BIC scoring criterion inherently penalizes model complexity by imposing a $\frac{1}{2} \ln N$ penalty per free parameter, which discourages overly dense structures. The learned BN structure is then fixed for all subsequent stages, so that the dependency-adjusted covariates become deterministic functions of the observed inputs.

In Stage II, the risk-weight parameters β are estimated by maximizing the improved group-wise partial likelihood, i.e.,

$$\mathcal{L}(\beta) = \sum_{g=1}^K -\frac{1}{N_g} \ln L_g(\beta), \quad (14)$$

where K is the total number of groups and N_g is the number of samples in group g . Here, $\mathcal{L}(\beta)$ denotes the weighted negative log-partial likelihood, to distinguish it from the partial likelihood $L(\beta)$ defined in (6). Notably, the partial likelihood of the g -th group can be derived from (6) as

$$L_g(\beta) = \prod_{i \in D_g(\tau)} \frac{\exp(f(\mathbf{x}_g^{(i)}; \beta))}{\sum_{j \in R_g(\tau)} \exp(f(\mathbf{x}_g^{(j)}; \beta))}, \quad (15)$$

where $R_g(\tau)$ denotes the risk set at time τ in group g , $D_g(\tau)$ denotes the set of failure events at time τ in group g , and f the risk function in (8). Due to its nonlinearity, the particle swarm optimizer (PSO) is employed to estimate $\hat{\beta}$ [33]. PSO is adopted because the BN-modulated risk function introduces non-convexity into the objective landscape, rendering gradient-based solvers unreliable.

In Stage III, the group-specific baseline cumulative hazard function is estimated using the Breslow estimator [29]:

$$\hat{H}_{0,g}(t) = \sum_{i \in D_g(\tau), \tau \leq t} \frac{d_\tau}{\sum_{j \in R_g(\tau)} \exp(f(\mathbf{x}_g^{(j)}; \hat{\beta}))}, \quad (16)$$

where d_τ denotes the number of failure events observed at time τ . The corresponding baseline hazard function is then obtained by discrete differentiation as

$$\hat{h}_{0,g}(t) = \frac{\Delta \hat{H}_{0,g}(t)}{\Delta t}. \quad (17)$$

Weighted Negative Log-partial Likelihood: In Stage II, the robustness of (14) is analyzed via the Fisher information and the Cramér–Rao lower bound (CRLB) [34]. Under the assumption that, conditional on group membership, the survival times of individual smart meters are independent and identically distributed [35], they are defined as

$$\begin{cases} I_g(\beta) = -\mathbf{E} \left(\frac{d^2}{d\beta^2} \ln L_g(\beta) \right), \\ \text{Var}(\hat{\beta}) \geq I(\beta)^{-1}, \end{cases} \quad (18)$$

where $\hat{\beta}$ denotes the estimated parameter vector. By substituting (15) into (18), it can be seen that $I_g(\beta)$ is positively related to the group sample size N_g . As a result, the total Fisher information $I(\beta) = \sum_g I_g(\beta)$ is mainly dominated by groups with large N_g , and the corresponding parameter estimation is also driven primarily by these groups according to the CRLB. However, for a model intended to generalize across heterogeneous groups, the contribution of small groups should not be neglected.

To reduce this imbalance, $w_g = 1/N_g$ is introduced into the negative log-partial likelihood, and the effective Fisher information of group g is

$$I_g^w(\beta) = w_g I_g(\beta), \quad (19)$$

which helps balance the contributions of different groups. Consequently, the weighted objective (14) improves the stability of parameter estimation for small groups and enhances the generalization ability of the proposed framework.

E. Failure Rate Prediction

For a given smart meter group, the total number of failures can be modeled as the sum of independent individual failure events, where the failure of each meter follows a Bernoulli distribution with specialized failure probability [36]. The probability mass function of the total number of failures K is

$$P(K = k) = \sum_{A \in F_k} \left[\prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j) \right], \quad (20)$$

where p_i is the failure probability of the i -th smart meter, F_k denotes the collection of all subsets of cardinality k , A denotes a subset of failed meters, and A^c denotes the complementary subset of meters that remain operational.

Accordingly, the expected group-level failure probability can be derived from the expectation of K as

$$F = \frac{1}{N} \mathbf{E}(K) = \frac{1}{N} \sum_{i=1}^N p_i, \quad (21)$$

where N is the total number of smart meters in the group. Therefore, the group-level failure probability is equal to the average of the individual failure probabilities within the group.

By further incorporating the survival function (12)-(13) of each smart meter, the time-dependent group-level failure probability can be expressed as

$$F(t) = \frac{1}{N} \sum_{i=1}^N (1 - S_i(t)) = 1 - \frac{1}{N} \sum_{i=1}^N \exp(-H_i(t)), \quad (22)$$

where $S_i(t)$ and $H_i(t)$ are the survival function and cumulative hazard function of the i -th smart meter, respectively. This provides a direct link between individual-level risk estimation and group-level prediction of failure probability.

F. Interpretability Analysis

Beyond prediction accuracy, the proposed interpretable failure-risk analysis mechanism provides explanations at two complementary levels, i.e., inter- and intra-group covariate effects. The p covariates are partitioned into an inter-group subset (indexed by $\mathcal{J}_{\text{inter}}$) and an intra-group subset (indexed by $\mathcal{J}_{\text{intra}}$). Accordingly, the risk function in (8) is decomposed as

$$f(\mathbf{x}; \boldsymbol{\beta}) = f_{\text{inter}}(\mathbf{x}; \boldsymbol{\beta}) + f_{\text{intra}}(\mathbf{x}; \boldsymbol{\beta}),$$

where

$$f_{\text{inter}}(\mathbf{x}; \boldsymbol{\beta}) = \sum_{j \in \mathcal{J}_{\text{inter}}} \beta_j x_j P(X_j = x_j | \text{Pa}(X_j)), \quad (23)$$

$$f_{\text{intra}}(\mathbf{x}; \boldsymbol{\beta}) = \sum_{j \in \mathcal{J}_{\text{intra}}} \beta_j x_j P(X_j = x_j | \text{Pa}(X_j)). \quad (24)$$

For any smart meter i in group g , its instantaneous hazard can be decomposed into two multiplicative components as

$$h_g(t | \mathbf{x}^{(i)}) = h_{0,g}(t) \exp(f_{\text{inter}}(\mathbf{x}^{(i)}; \boldsymbol{\beta})) \exp(f_{\text{intra}}(\mathbf{x}^{(i)}; \boldsymbol{\beta})), \quad (25)$$

where $h_{0,g}(t) \exp(f_{\text{inter}})$ represents the group-level failure risk determined by the baseline hazard and inter-group covariates, and $\exp(f_{\text{intra}})$ captures individual-level variations

within the group. The resulting decomposition enables direct quantification of covariate contributions to failure risk in a closed-form and interpretable manner.

1) Inter-group Covariate Effects: Inter-group covariates (e.g., Production Batch, Distribution Area, and Battery Selection) remain constant within each group. Their effects are reflected through both the group hazard ratio $\exp(f_{\text{inter}})$ and the group-specific baseline hazard $h_{0,g}(t)$. Since these covariates do not vary among meters within the same group, $\exp(f_{\text{inter}})$ only provides a constant proportional adjustment to the hazard. The temporal evolution of failure risk, including the location and magnitude of hazard peaks, is therefore primarily encoded in the baseline hazard function $h_{0,g}(t)$.

To identify the inter-group covariates associated with observed hazard peaks, a posterior attribution method based on Bayes' theorem is employed. Let X_j ($j \in \mathcal{J}_{\text{inter}}$) denote an inter-group covariate, and let $(t_{\text{pk}}, a_{\text{pk}})$ denote the time and area of a detected baseline hazard peak. The posterior probability that the peak is associated with covariate value x is computed as

$$P(X_j = x | t_{\text{pk}}, a_{\text{pk}}) = \frac{P(t_{\text{pk}}, a_{\text{pk}} | X_j = x) P(X_j = x)}{\sum_{x'} P(t_{\text{pk}}, a_{\text{pk}} | X_j = x') P(X_j = x')}, \quad (26)$$

where the prior probability $P(X_j = x)$ is estimated from the sample-weighted frequency of groups with $X_j = x$, and the likelihood is estimated using a two-dimensional kernel density estimator fitted to the empirical peak distribution.

The dominant covariate value for each peak is determined by the maximum posterior probability. An area-weighted attribution score is then computed and subsequently normalized to quantify the overall explanatory capability of each inter-group covariate. The resulting score $\mathcal{E}(X_j)$ represents the proportion of total explanatory power attributed to covariate X_j .

Together, the group hazard ratio and the posterior-based attribution analysis provide complementary interpretations of inter-group effects. The former quantifies the overall shift in failure risk, whereas the latter identifies the dominant covariate values associated with specific temporal failure patterns.

2) Intra-group Covariate Effects: Intra-group covariates (e.g., Measurement Error, Clock Error, and Inventory Time) vary among individual meters within the same group. Their effects are captured by the within-group hazard ratio $\exp(f_{\text{intra}})$, which describes the deviation of an individual meter's failure risk from the group baseline.

Since the BN-enhanced risk function is additive, each term

$$\beta_j x_j P(X_j = x_j | \text{Pa}(X_j)) \quad (27)$$

can be interpreted independently. A positive contribution increases the hazard, whereas a negative contribution reduces it. This formulation enables direct quantification of the contribution of each intra-group covariate to individual failure risk.

IV. RESULTS AND DISCUSSIONS

A. Experimental Settings

We present the detailed experimental settings used to validate the effectiveness of the proposed framework.

1) *Dataset preparation*: The training and test sets are sampled from the full dataset, which contains survival time and 11 covariates. A stratified random sampling strategy with a ratio of 7:3 is adopted to preserve the distributions of failure events and covariate characteristics across groups, as recommended in [37]. The covariates are then classified into two categories: inter-group and intra-group. The inter-group covariates are categorical variables, including Installation City, Oscillator Selection, Production Batch, Battery Selection, Manufacturers, Distribution Area, and MCU Selection. These variables vary across groups but remain constant within each group. In contrast, the intra-group covariates are continuous variables, including Measurement Error, Clock Error, Inventory Time, and Inventory Time Ratio, which vary both across groups and among meters within the same group.

2) *Compared models*: Comparison experiments are conducted using different datasets and variants. To evaluate the advantage of the proposed semi-parametric framework over parametric alternatives, the Weibull and Gaussian models are included for comparison. In addition, an ablation study is designed to examine the effectiveness of the proposed grouping strategy and the BN-enhanced risk formulation. The compared models are summarized as follows:

- **Classical CPHM**: the conventional survival analysis model, which serves as a baseline and estimates regression coefficients through likelihood-based fitting [38].
- **Bayesian linear estimation CPHM**: a CPHM variant that introduces Bayesian linear regression to quantify parameter uncertainty and improve estimation stability, especially under small-sample conditions [39].
- **BN-interpreted CPHM**: the mode preserves the linear covariate–hazard relationship of the classical CPHM while incorporating a BN to describe probabilistic dependencies between covariates and failure events [40].
- **Group-wise single-BN CPHM**: a variant where each group is assigned an individual BN combined with its corresponding CPHM, rather than using a shared global BN, to model within-group covariate dependencies.

3) *Performance metrics*: Since the failure events are sparse, the failure rate estimated directly from observed failures is sensitive to noise. Therefore, the failure probability is adopted as the primary criterion for performance evaluation and model comparison. Accordingly, the concordance index (C-index) and the coefficient of determination R^2 are adopted as the evaluation metrics.

The C-index measures the model's ability to correctly rank survival times, and its value ranges from 0.5 to 1. A larger C-index indicates stronger discriminative capability in identifying relative failure risk [41]. Here, the C-index mainly reflects the effectiveness of the learned BN structure and the corresponding risk-score function in distinguishing smart meters with different failure tendencies. Therefore, it assesses the ranking performance of the model rather than the direct accuracy of failure rate or failure probability prediction.

To further assess the quantitative accuracy of failure prediction, the mean R^2 values between the predicted failure probabilities and the observed failure ratios at failure-event

times are calculated for each group. Compared with the C-index, the R^2 metric more directly evaluates how accurately the model reproduces the evolution of failure probability within each group. Moreover, we report the mean AUC, defined as the average of time-dependent $AUC(t)$ [42] over all observed failure times. $AUC(t)$ quantifies the model's ability to discriminate between smart meters that fail before time t and those that survive beyond t ; a value of 0.5 indicates random prediction, while values above 0.7 suggest useful discriminative capability. Unlike the single-threshold AUC used in standard classification [43], the mean AUC here is inherently time-dependent and summarizes discriminative performance over the entire observation period.

B. Model Visualization and Analysis

1) *BN Structure and Covariate Importance Analysis*: The learned BN structure is shown in Figs. 2, which presents the dependency structure among manufacturing, environmental, and operational factors. For example, the distribution area is a parent node of the installation city, indicating that the installation location is conditionally dependent on the deployment region. Overall, the BN organizes the main influencing factors into four categories, namely manufacturing, component, environmental, and technical performance factors, and provides an interpretable description of their dependency relationships.

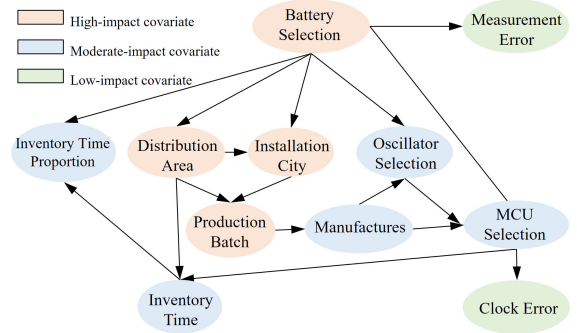


Fig. 2: Structure of the integrated Bayesian network, where the nodes represent covariates, and the directed edges represent dependencies among covariates.

Figure 3 further shows the estimated risk weights of all covariates, which reflect both the direction and relative magnitude of their effects on failure risk. The classification is based on the maximum absolute risk-weight value across all categories for each covariate, $|\beta_j|_{\max} = \max_k |\beta_{j,k}|$, which quantifies the maximum possible deviation in risk score attributable to that covariate. According to $|\beta_j|_{\max}$, the covariates can be broadly categorized into three levels of influence:

- **High-impact covariates**: Distribution Area, Production Batch, Installation City, and Battery Selection. These inter-group covariates reflect differences in operating environment, manufacturing quality, and power supply stability, which directly affect long-term failure risk.
- **Moderate-impact covariates**: MCU Selection, Oscillator Selection, Manufacturers, Inventory Time, and Inventory Time Ratio. These covariates capture component

characteristics and storage conditions, e.g., oscillator deviations may cause long-term clock drift, and extended inventory time may worsen pre-installation degradation.

- **Low-impact covariates:** Measurement Error and Clock Error during initial verification, which are tightly controlled during verification and therefore vary within narrow ranges.

Overall, the strongest effects are primarily associated with inter-group covariates, indicating that smart meter failure behavior is strongly affected by group-level heterogeneity. If grouping were not introduced, model fitting would be dominated by inter-group differences, making it difficult to describe fine-grained failure patterns within individual groups. Therefore, the covariate importance analysis further confirms the necessity and validity of the proposed grouping strategy.

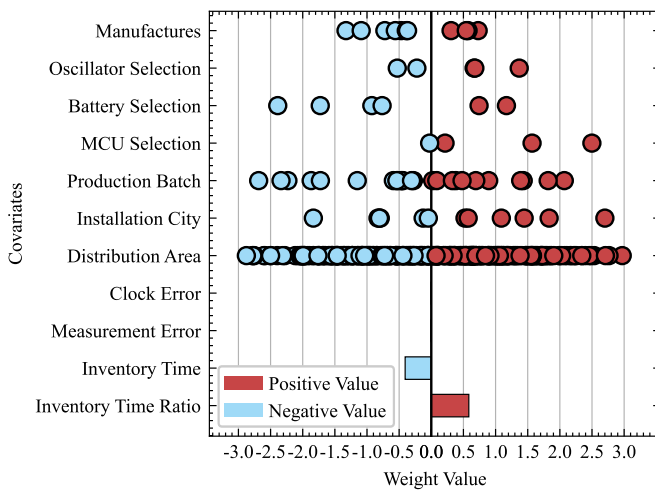


Fig. 3: Risk weights of all covariates. Positive values indicate positive contributions to failure risk, whereas negative values indicate risk-reducing effects.

2) *Failure Rate & Failure Probability Prediction:* Since failure events are sparse within each group, the evaluation is primarily conducted using failure probability curves derived from the predicted hazards, together with empirical failure ratios. Note that direct estimation of failure rates from observed failures is unstable and can be distorted by numerical differencing under sparse events; for this reason, the failure probability is adopted as the primary evaluation criterion. For each smart meter, the failure rate and failure probability are computed from the BN-enhanced risk score and the corresponding group-specific baseline hazard, and the group-level behavior is obtained by aggregating individual predictions.

The failure probability curves of different groups are given in Fig. 4. Groups within the same production batch show similar trends, including comparable onset times of nonzero failure probability and similar growth patterns. Meanwhile, noticeable differences remain across distribution areas. This result indicates that smart meters within the same batch share common reliability characteristics, yet exhibit region-dependent heterogeneity.

To further evaluate the prediction performance, three representative groups are selected for detailed analysis, as summa-

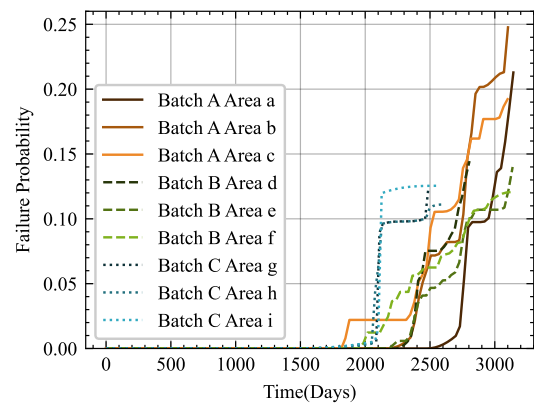


Fig. 4: Comparison of failure probability curves across groups under different production batches and distribution areas.

rized in Table I. These groups differ in the number of smart meters, the number of failure events, and the censoring ratio, and therefore provide representative test cases under different data conditions.

TABLE I: Comparison of Three Selected Groups in terms of Number of Smart Meters, Number of Failure Events, and Censoring Ratio.

Dataset	Group	Total Number	Failure Events	Censoring Ratio
Training	1	71	11	0.84
	2	168	19	0.89
	3	71	24	0.66
Test	1	31	5	0.84
	2	72	9	0.88
	3	30	11	0.63

The detailed baseline hazard curves and corresponding failure probability curves are displayed in Fig. 5. For all three groups, the predicted failure probability remains close to zero during the early operating period, approximately from 0 to 2000 days. This is consistent with the very low baseline hazards observed in the corresponding subplots, indicating that the cumulative failure risk is negligible in the early life stage of smart meters. After approximately 2000 days, all three groups exhibit a rapid increase in failure probability, coinciding with pronounced peaks in the baseline hazard curves. This result indicates that the growth in cumulative failure probability is driven directly by increases in instantaneous failure risk. It also suggests that the failure behavior of smart meters does not necessarily follow a classical bathtub-shaped curve, but may instead exhibit characteristics with repeated hazard peaks. A possible reason is that smart meters are subject to multiple failure mechanisms with distinct onset times: early peaks may be associated with manufacturing defects or component quality variations, whereas later peaks may reflect cumulative degradation driven by environmental stress and aging. The coexistence of these mechanisms produces a superposition of hazard contributions that cannot be captured by a single-mode parametric model, further motivating the use of a nonparametric baseline hazard formulation as adopted in the proposed framework.

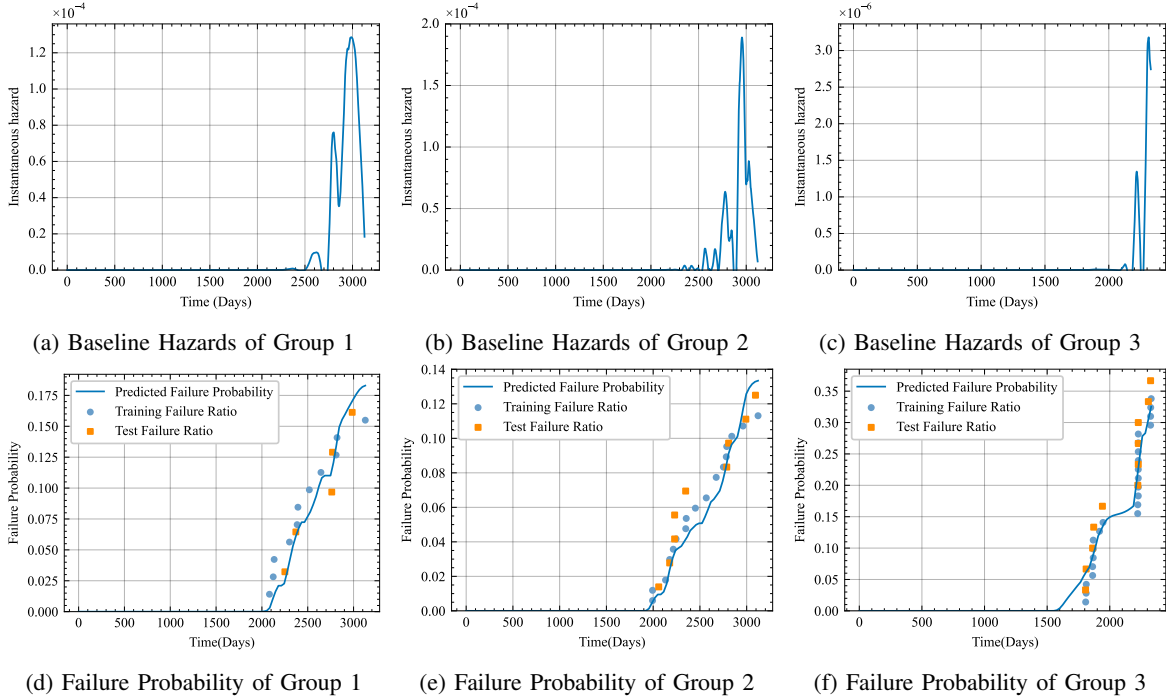


Fig. 5: Failure probability curves (a–c) for three representative groups, together with empirical failure ratios.

Moreover, in all three groups, the predicted failure probability curves closely follow the empirical failure ratios in both the training and test sets. This indicates that the global–local modeling architecture can accurately capture the main failure dynamics of smart meters and that the identified hazard peaks correspond to actual increases in field failure risk.

3) *Model Assessment*: Here we quantify the performance of the proposed model in terms of C-index and R^2 . For each group, the data are split into training and test subsets, and the C-index and R^2 are computed separately on each subset. The proposed model achieves consistently high C-index (0.95/0.95) and R^2 (0.94/0.90) values on both the training and test sets. These results indicate that the model not only provides reliable survival-risk ranking, but also accurately captures the failure-probability trends of smart meter groups. Overall, the evaluation results demonstrate the effectiveness of the proposed BN-enhanced CPHM framework in both risk discrimination and failure prediction.

4) *Interpretability Case Study*: To validate the proposed interpretable failure-risk analysis mechanism, a representative failure sample from Group 3 is selected to demonstrate the three-part hazard decomposition defined in (25). This meter experienced failure at $t = 2228$ days; the group-specific baseline hazard $h_{0,g}(t)$ is non-zero in $[1808, 2331]$ days with cumulative hazard $H_{0,g}(t_{\max}) = 2.31 \times 10^{-4}$, estimated from 24 observed event times.

- **Inter-group HR**. Table II lists the per-variable hazard ratio $HR_j = \exp(\beta_j P(X_j = x_j | \text{Pa}(X_j)))$ for each inter-group covariate, computed via (23). The inter-group covariates yield $HR_{\text{group}} = 0.423$, reducing the hazard to 42.3% of the baseline level. Battery Selection ($HR_j = 0.451$) is the dominant risk-reducing

TABLE II: Inter-group covariate contributions for the representative sample

Variable	β_j	$P(X_j \text{Pa})$	HR_j
Battery Selection	−2.389	0.333	0.451
MCU Selection	+1.568	0.360	1.759
Manufacturers	−0.389	1.000	0.678
Oscillator Selection	−0.529	0.249	0.877
Production Batch	−1.725	0.057	0.906
Installation City	−0.127	0.112	0.986
Distribution Area	+0.200	0.017	1.003
$HR_{\text{inter-group}} = \prod HR_j$			0.423

factor, while MCU Selection ($HR_j = 1.759$) is the only risk-increasing covariate; Manufacturers ($HR_j = 0.678$) provides a secondary protective contribution.

- **Intra-group HR**. Table III lists the per-variable hazard ratio $HR_j = \exp(\beta_j x_j P(X_j = x_j | \text{Pa}(X_j)))$ for each intra-group covariate, computed via (24).

TABLE III: Intra-group covariate contributions for the representative sample

Variable	β_j	x_j	$P(X_j \text{Pa})$	HR_j
Inventory Time	−0.407	3.191	0.012	0.984
Inventory Time Ratio	+0.584	−4.518	0.013	0.966
Measurement Error	−0.001	0.056	0.025	1.000
Clock Error	+0.000	−0.022	0.009	1.000
$HR_{\text{intra-group}} = \prod HR_j$				0.951

The intra-group covariates yield $HR_{\text{within}} = 0.951$, providing a modest individual-level protective effect. Inventory Time Ratio ($HR_j = 0.966$) and Inventory Time ($HR_j = 0.984$) are the main contributors, while Measure-

ment Error and Clock Error have negligible impact. The total hazard ratio is $HR_{\text{total}} = HR_{\text{group}} \times HR_{\text{within}} = 0.423 \times 0.951 = 0.402$.

- **Baseline hazard attribution $\mathcal{E}(X_j)$.** Table IV reports the normalized explanatory power $\mathcal{E}(X_j)$ for each inter-group covariate across all detected baseline hazard peaks. Each value denotes the proportion of total explanatory power attributable to that covariate, with all seven values summing to 1.

TABLE IV: Normalized baseline hazard peak attribution

Covariate	$\mathcal{E}(X_j)$
MCU Selection	0.197
Oscillator Selection	0.189
Installation City	0.151
Manufacturers	0.122
Battery Selection	0.120
Distribution Area	0.113
Production Batch	0.109

MCU Selection ($\mathcal{E}=0.197$) and Oscillator Selection ($\mathcal{E}=0.189$) together account for 38.6% of the total explanatory power, confirming that component selection is the primary driver of baseline hazard peak shape. Installation City ($\mathcal{E}=0.151$) ranks third, reflecting the influence of the deployment environment. The remaining four covariates each contribute approximately 11–12%, indicating that every inter-group covariate plays a non-negligible role in differentiating the temporal failure pattern.

In summary, the three-part decomposition reveals: (i) the baseline hazard $h_{0,g}(t)$ determines the temporal risk window, with MCU Selection ($\mathcal{E}=0.197$) and Oscillator Selection ($\mathcal{E}=0.189$) contributing the most to the hazard-peak shape; (ii) the inter-group HR ($=0.423$) provides the dominant proportional adjustment driven by Battery Selection and MCU Selection; and (iii) the intra-group HR ($=0.951$) adds a modest individual-level protective effect, yielding $HR_{\text{total}}=0.402$.

C. Comparison with Different Methods

To further evaluate the proposed BN-enhanced CPHM, comparison experiments are conducted with respect to prediction accuracy, generalization ability, and adaptability to grouped data. For fair comparison, the same dataset and representative groups as those in Section IV-B2 are used for visualization and quantitative evaluation. The compared methods are organized into three categories, which are presented in the following part.

1) *Comparison with Prediction Models of Different Distributions:* To examine the advantage of the proposed semi-parametric formulation, two parametric models with different distributional assumptions are considered: the Weibull and Gaussian models. Both models assume that the hazard function follows a specific parametric form, a common assumption in traditional survival modeling.

The results are summarized in Table V. The Weibull and Gaussian models achieve C-index values of 0.67/0.66 and 0.85/0.85 on the training/test sets, respectively, both of which remain inferior to the proposed BN-enhanced CPHM, which

TABLE V: Mean Performance Metrics of Prediction Models of Different Distributions

Metric	Grouping Strategy	Train Set	Test Set
C-Index	Weibull Model	0.67	0.66
	Gaussian Model	0.85	0.85
	Proposed Model	0.95	0.95
R^2	Weibull Model	<0.5	<0.5
	Gaussian Model	<0.5	<0.5
	Proposed Model	0.94	0.90
AUC	Weibull Model	0.69	0.69
	Gaussian Model	0.69	0.69
	Proposed Model	0.97	0.97

achieves 0.95 on both sets. Though the Gaussian model performs better than the Weibull model, both remain clearly inferior to the proposed BN-enhanced CPHM, which achieves a C-index of 0.95 on both the training and test sets. A similar trend is observed for R^2 , where both parametric models remain below 0.5, indicating limited ability to fit the observed failure probability curves. In terms of mean AUC, both the Weibull and Gaussian models achieve only 0.69, compared with 0.97 for the proposed model, indicating that the parametric models fail to provide useful time-dependent discriminative capability across the observation period.

The visual comparisons in Fig. 6 are consistent with these results: the Weibull and Gaussian models deviate noticeably from the empirical failure ratios, whereas the proposed model follows the observed trends more closely.

The inferior performance of the two parametric models is mainly due to their rigid distributional assumptions on the hazard function, which limit flexibility when the actual failure process deviates from the assumed form. Specifically, the Weibull model assumes a monotonic hazard that cannot capture multimodal peaks, while the Gaussian model imposes a symmetric distribution inconsistent with the right-skewed failure behavior of electronic devices.

Under sparse failure observations, these fixed assumptions also make the fitted curves more sensitive to the limited number of failures and less effective in utilizing censored survival information. In contrast, the proposed BN-enhanced CPHM does not impose a fixed parametric form on the baseline hazard and fully incorporates censored data. By combining BN-based dependency modeling with the survival-analysis capability of the CPHM, it provides a more flexible and accurate description of smart meter failure behavior.

2) *Comparison with Prediction Models without Grouping:* To examine the effect of the proposed grouping strategy, the BN-enhanced CPHM is compared with a non-grouping variant. The comparison focuses on both prediction accuracy and generalization performance. The mean evaluation metrics are reported in Table VI, and the corresponding failure probability curves for three representative groups are shown in Fig. 7.

The obtained results in Table VI show that the grouped model substantially outperforms the non-grouping model. The proposed model improves the C-index from 0.63/0.64 to 0.95/0.95 on the training/test sets and increases the R^2 from 0.76/0.75 to 0.94/0.90. The mean AUC further confirms this

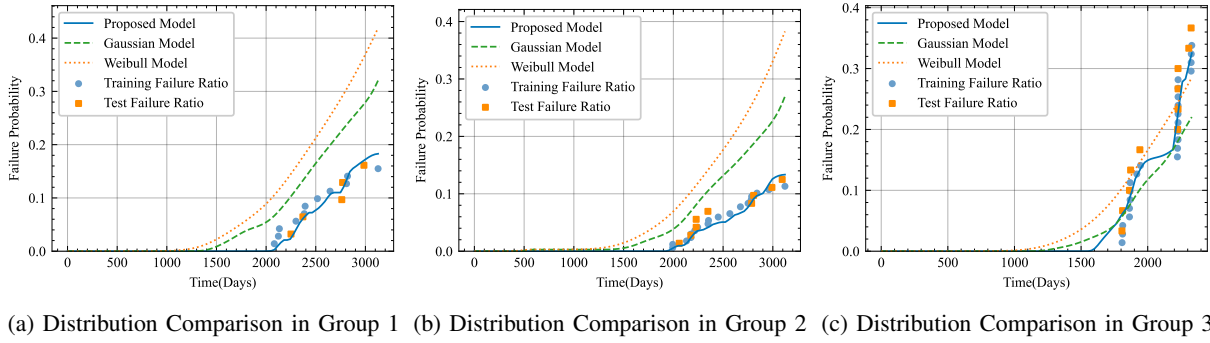


Fig. 6: Failure probability curve comparison among different distribution-based models for three representative groups.

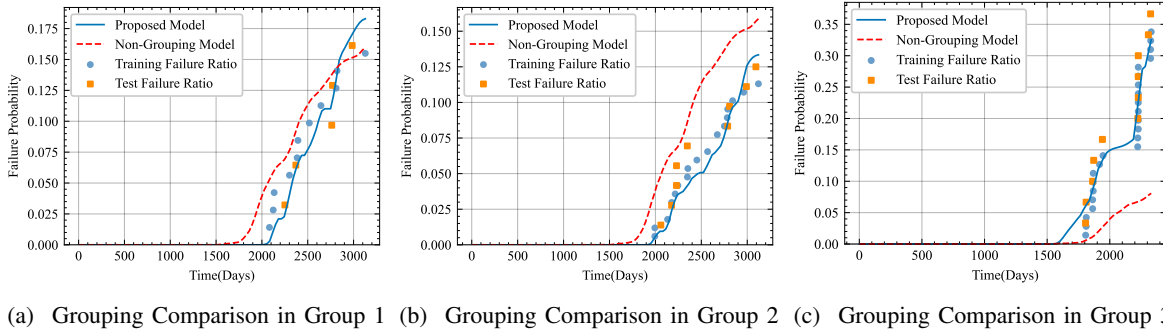


Fig. 7: Failure probability curve comparison between grouping and non-grouping strategies for three representative groups.

TABLE VI: Mean Performance Metrics of Prediction Models with and without Grouping

Metric	Grouping Strategy	Train Set	Test Set
C-Index	Non-Grouping Model	0.63	0.64
	Proposed Model	0.95	0.95
R ²	Non-Grouping Model	0.76	0.75
	Proposed Model	0.94	0.90
AUC	Non-Grouping Model	0.56	0.53
	Proposed Model	0.97	0.97

gap: the non-grouping model achieves only 0.56/0.53, barely above random (0.5), whereas the proposed model reaches 0.97/0.97, demonstrating consistently strong time-dependent discrimination when grouping is applied. It is worth noting that the C-index improvement is particularly significant: the non-grouping model achieves only 0.63/0.64, which is close to random ranking (0.5), indicating that without grouping the model cannot effectively distinguish relative failure risk among individual smart meters. These improvements indicate that the grouping strategy is essential for both survival-risk ranking and failure probability prediction.

The visualization in Fig. 7 further illustrates this advantage. The grouped model produces failure probability curves that more closely match the empirical failure ratios across all three representative groups, whereas the non-grouping model exhibits larger deviations from the observed trends. This result indicates that explicit grouping provides a more accurate description of group-specific failure behavior.

The inferior performance of the non-grouping model can be

attributed to its use of a single global baseline hazard function across all smart meters. Under this setting, the model is forced to fit an average failure mode across all samples. Although the risk scores may still vary among groups, the shared baseline hazard masks differences in failure dynamics across production batches and deployment areas. More specifically, different groups exhibit distinct hazard peak locations and magnitudes (as shown in Fig. 5), reflecting group-specific failure mechanisms driven by manufacturing and environmental factors. A single global baseline hazard cannot represent these heterogeneous temporal patterns simultaneously; it smooths out group-specific peaks and troughs, resulting in a compromise curve that accurately describes no individual group. Consequently, the model tends to underestimate failure risk in high-risk groups and overestimate it in low-risk groups, thereby weakening its overall generalization ability.

Furthermore, the non-grouping model has a noticeably higher R^2 (0.76/0.75) than C-index (0.63/0.64), suggesting that while the global baseline can partially track the average trend of failure probability accumulation, it fails to capture the relative risk ordering among individual meters—a task that requires group-specific hazard differentiation.

In contrast, the proposed grouping strategy allows different groups to have their own baseline hazard functions while sharing the BN-enhanced risk formulation. This global-local architecture preserves common dependency information across groups through the shared BN, while capturing group-specific failure dynamics through individual baseline hazards. Therefore, the comparison results clearly demonstrate the necessity of the proposed grouping strategy.

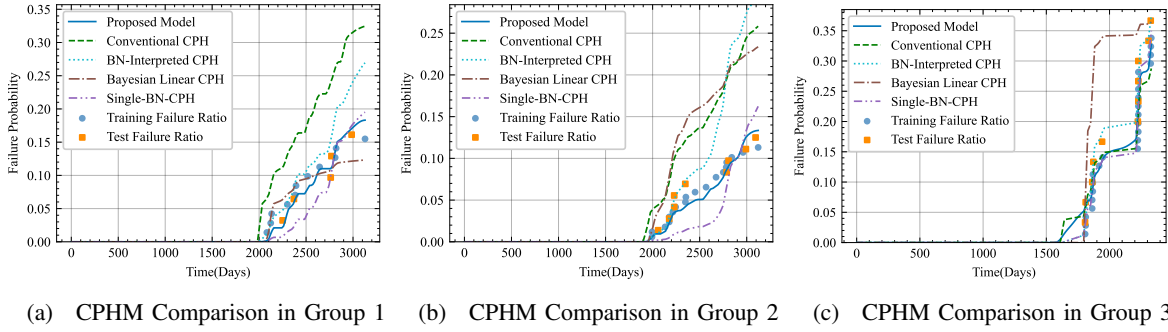


Fig. 8: Failure probability curve comparison among different CPHM-based variants in three representative groups.

3) *Comparison with CPHM Variants*: To further evaluate the effectiveness of the proposed BN-enhanced CPHM, it is compared with several representative CPHM variants, including the classical CPHM, the Bayesian linear CPHM, the BN-interpreted CPHM, and the single-BN CPHM, as mentioned in Section IV-A. They are compared to assess the effects of different approaches to integrating probabilistic modeling with the Cox framework.

TABLE VII: Mean Performance Metrics of Different CPHM-Based Variants

Metric	Types of CPHM	Train Set	Test Set
C-Index	Conventional	0.95	0.75
	Bayesian Linear	0.90	0.87
	BN Interpreted	0.91	0.82
	Single BN	0.90	0.80
	Proposed Model	0.95	0.95
R^2	Conventional	<0.5	<0.5
	Bayesian Linear	<0.5	<0.5
	BN Interpreted	0.65	0.59
	Single BN	0.71	0.65
	Proposed Model	0.94	0.90
AUC	Conventional	1.00	1.00
	Bayesian Linear	0.95	0.95
	BN Interpreted	0.95	0.95
	Single BN	0.50	0.50
	Proposed Model	0.97	0.97

The results are summarized in Table VII, and the corresponding curve comparisons are shown in Fig. 8. Overall, the proposed BN-enhanced CPHM achieves the best performance among all compared variants, with a C-index of 0.95 on the training and test sets and R^2 values of 0.94 and 0.90, respectively. In contrast, the conventional CPHM exhibits a clear drop in test-set C-index and a very low R^2 , indicating limited generalization and weak curve-fitting performance.

The conventional CPHM achieves a high training C-index but suffers from severe overfitting due to the lack of dependency-aware regularization. The Bayesian linear CPHM stabilizes estimation through prior distributions but does not model inter-covariate dependencies, resulting in suboptimal R^2 . The BN-interpreted and single-BN CPHMs further improve R^2 by incorporating dependency information; however, the former lacks grouping to capture inter-group heterogeneity, and the latter fragments learning across small within-group samples. Neither variant matches the proposed model in both

ranking accuracy and failure probability fitting. Regarding mean AUC, the conventional CPHM achieves 1.00 (overfitting), the Bayesian linear and BN-interpreted CPHMs reach 0.95, and the single-BN CPHM drops to 0.50; the proposed model attains 0.97, demonstrating robust time-dependent discrimination.

These results suggest that the advantage of the proposed model lies in the joint use of a shared global BN and group-specific baseline hazard functions. The global BN captures dependencies among covariates across all groups, improving robustness under high censoring and sparse failures, whereas the group-specific baseline hazards adapt to inter-group heterogeneity in failure dynamics. Their combination enables the proposed framework to achieve both accurate fitting and strong generalization across different smart meter groups.

D. Real-world Validation

To verify the practicality of the proposed framework, a real-world dataset is collected from a new deployment area. This dataset contains 840 smart meters from 7 groups, including meters that are still in operation. The proposed BN-enhanced CPHM is applied to this dataset to predict the failure rates and failure probabilities of the smart meters. The model achieves C-index values of 0.96 and 0.95 on the training and test sets, respectively, indicating strong ranking performance on the new dataset. The corresponding R^2 values are 0.91 and 0.82, showing that the model also maintains satisfactory accuracy in failure-probability prediction under real-world conditions. The mean AUC on the validation dataset is 0.9878, further confirming the model's strong time-dependent discriminative capability on unseen deployment data.

The distribution of prediction residuals is presented in Fig. 9. The residuals are centered near zero within a relatively narrow range, indicating that the predicted failure probabilities are generally consistent with the observed outcomes.

To further assess the calibration of predicted failure probabilities, a calibration plot is constructed by comparing the predicted failure probabilities against the observed failure ratio across decile-based risk groups, as shown in Fig. 10. Each point represents one risk group, and the diagonal line indicates perfect calibration. The close alignment between the predicted and observed values confirms that the proposed model produces well-calibrated probability estimates, without systematic over- or under-prediction.

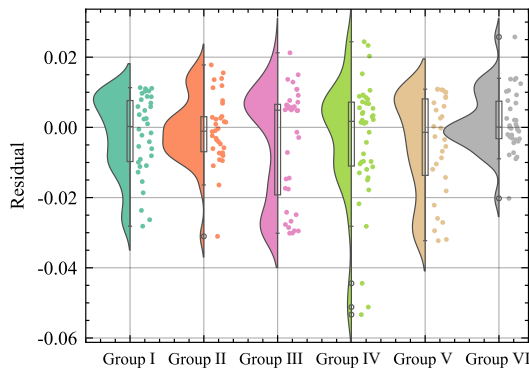


Fig. 9: Distribution of failure prediction residuals for the real-world dataset, where the residual means are close to 0, and the ranges are within $[-0.06, 0.04]$.

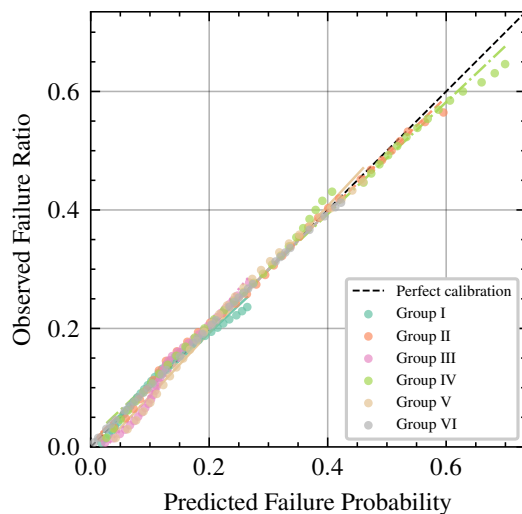


Fig. 10: Calibration plot comparing predicted failure probabilities against observed failure ratios across decile-based risk groups. The dashed diagonal indicates perfect calibration.

The above results indicate that the global-local modeling architecture generalizes well to unseen field data and can provide effective support for maintenance scheduling and replacement planning.

V. CONCLUSIONS

This paper proposes a BN-enhanced CPHM for grouped smart meter failure-rate prediction. The results show that the grouping strategy based on production batch and deployment area is effective for characterizing heterogeneous failure behavior in smart meters. By introducing group-specific baseline hazards, the global-local modeling architecture captures inter-group differences, thereby improving the description of group-level failure dynamics and enhancing model generalization. The BN-enhanced CPHM also achieves a balance between predictive performance and interpretability. Specifically, the BN models dependencies among multiple covariates and provides interpretable risk weights for key influential factors, whereas the CPHM naturally handles right-censored data and uses a

nonparametric baseline hazard function to adapt to observed failure trends in smart meters. Furthermore, the interpretable failure-risk analysis mechanism, comprising the posterior-based peak attribution and single-sample hazard decomposition, provides transparent, closed-form explanations of failure risk at both the group and individual levels, substantiating the interpretability of the proposed framework.

Experiments on real operational data from 38,238 smart meters across 24 production batches and 358 deployment areas demonstrate the effectiveness of the proposed method. The model achieves a test-set C-index of 0.95 and an R^2 of 0.90, outperforming traditional parametric models, non-grouped models, and other CPHM-based variants. Additional validation on an independent real-world dataset further confirms its practical value for smart meter failure prediction and maintenance decision support.

Future work will focus on incorporating time-varying covariates to better characterize the dynamic evolution of failure risk and on exploring transfer learning strategies to improve predictive performance for groups with limited failure samples. In addition, the rapid advancement of large language models and multimodal learning paradigms in machine monitoring and fault diagnostics [44] suggests that integrating such emerging techniques with the proposed interpretable framework could further enhance diagnostic capability and adaptive maintenance decision support for smart metering systems.

REFERENCES

- [1] C.-C. Sun, D. J. Sebastian Cardenas, A. Hahn, and C.-C. Liu, "Intrusion Detection for Cybersecurity of Smart Meters," *IEEE Transactions on Smart Grid*, vol. 12, pp. 612–622, Jan. 2021.
- [2] J. C. Parath and A. K. Saha, "Smart Meter Data Analytics Using Decision Trees and Nearest-Neighbours," in *2023 IEEE AFRICON*, pp. 1–6, Sept. 2023.
- [3] D. Xu, X. Xiao, and Y. Haibo, "Reliability Evaluation of Smart Meters Under Degradation-Shock Loads Based on Phase-Type Distributions," *IEEE Access*, vol. 8, pp. 39734–39746, 2020.
- [4] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," *IEEE Transactions on Smart Grid*, vol. 10, pp. 3125–3148, May 2019.
- [5] N. Li, X. Dong, G. Chen, T. Yuan, and Z. Duan, "A smart meters optimization model for reliability prediction," in *2016 International Conference on Integrated Circuits and Microsystems (ICICM)*, pp. 159–165, Nov. 2016.
- [6] Y. Wang, N. Liu, J. Yang, Y. Dai, M. Wang, Y. Jia, G. Gu, X. Dai, B. Pang, and L. Gu, "Analysis on the reasons for smart meter failure and solutions," in *Xian, PEOPLES R CHINA proceedings of 2019 4th International Workshop on Materials Engineering and Computer Sciences (IWMECS 2019)*, pp. 475–479, 2019.
- [7] P. Wang, Z. Jing, X. Dong, L. Guo, Z. Wang, Z. Li, and K. Liu, "A reliability prediction method of smart meter based on dynamic bayesian networks," in *12th International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE 2022)*, pp. 968–972, 2022.
- [8] Z. Yang, Y. X. Chen, Y. F. Li, E. Zio, and R. Kang, "Smart electricity meter reliability prediction based on accelerated degradation testing and modeling," *International Journal of Electrical Power & Energy Systems*, vol. 56, pp. 209–219, Mar. 2014.
- [9] Y. Li, D. Qin, H. V. Poor, and Y. Wang, "Introducing edge intelligence to smart meters via federated split learning," *Nature Communications*, vol. 15, p. 9044, Oct. 2024.
- [10] K. Xia, Z. Zhou, and J. Wang, "Smart meter-based system for real-time fault diagnosis of inter-turn short circuits in induction motors leveraging attention-enhanced resnet," *IEEE Instrumentation & Measurement Magazine*, vol. 28, pp. 3–10, Nov. 2025.

- [11] Y. Lan, Y. Zhang, W. Yao, S. Wei, T. Chen, H. Shuai, and J. Wen, "Trustworthy data-driven prefault frequency security assessment for power systems using disturbance-informed learning," *IEEE Transactions on Industrial Informatics*, vol. 22, no. 3, pp. 1950–1961, 2026.
- [12] L. Chen, K.-W. Lao, Y. Ma, and Z. Zhang, "Error Modeling and Anomaly Detection of Smart Electricity Meter Using TSVD+L Method," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022.
- [13] Z. Zhao, Y. Chen, J. Liu, Y. Cheng, C. Tang, and C. Yao, "Evaluation of Operating State for Smart Electricity Meters Based on Transformer–Encoder–BiLSTM," *IEEE Transactions on Industrial Informatics*, vol. 19, pp. 2409–2420, Mar. 2023.
- [14] Y. Lan, S. Wei, W. Yao, Y. Zhang, Y. Yang, and J. Wen, "Reapplication of training set: A physically reliable framework for power systems dominant instability mode identification using similar samples," *IEEE Transactions on Power Systems*, vol. 41, no. 1, pp. 597–610, 2026.
- [15] H. Cai, H. Chen, X. Ye, X. Zhang, H. Wen, J. Li, and Q. Guo, "An On-Line State Evaluation Method of Smart Meters Based on Information Fusion," *IEEE Access*, vol. 7, pp. 163665–163676, 2019.
- [16] C. Chen, Y. Liu, S. Wang, X. Sun, C. Di Cairano-Gilfedder, S. Titmus, and A. A. Syntetos, "Predictive maintenance using Cox proportional hazard deep learning," *Advanced Engineering Informatics*, vol. 44, p. 101054, 2020.
- [17] Y. Gao, S. Li, D. Wang, J. Mao, and L. Ouyang, "A neural network-based survival analysis model considering censored data for failure prediction," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 24585–24598, 2025.
- [18] B. Cui, S. Ahsan, J. Zhang, and X. Liang, "An improved cox proportional hazards model for reliability analysis of aviation gas turbines considering varying environmental conditions and operational settings," *Reliability Engineering and System Safety*, vol. 265, p. 111451, 2026.
- [19] N. Hagemann, T. Kneib, and K. Möllenhoff, "Capturing heterogeneous time-variation in covariate effects in non-proportional hazard regression models," *arXiv preprint arXiv:2501.13525*, 2025.
- [20] X. Yu, Z. Xu, X. Zhou, J. Zheng, Y. Xia, L. Lin, and S.-H. Fang, "Load Forecasting Based on Smart Meter Data and Gradient Boosting Decision Tree," in *2019 Chinese Automation Congress (CAC)*, pp. 4438–4442, Nov. 2019.
- [21] X. Kong, X. Zhang, N. Lu, Y. Ma, and Y. Li, "Online Smart Meter Measurement Error Estimation Based on EKF and LMRLS Method," *IEEE Transactions on Smart Grid*, vol. 12, pp. 4269–4279, Sept. 2021.
- [22] L. Ma, Z. Meng, Z. Teng, and W. Qiu, "A reliability evaluation framework for smart meters based on AGG-ARIMA and PFR," *Measurement Science and Technology*, vol. 33, p. 045006, Jan. 2022.
- [23] F. E. AbuAladas, M. Shehab, R. Israwah, G. Jaradat, and Y. Qawqzeh, "A new hybrid approach for improving intrusion detection system based on scatter search algorithm and support vector machine," *Journal of Computational and Cognitive Engineering*, vol. 5, no. 1, pp. 143–151, 2026.
- [24] Y. Xie and S. P. Tummala, "Machine learning for sensor analytics: A comprehensive review and benchmark of boosting algorithms in healthcare, environmental, and energy applications," *Sensors (Basel, Switzerland)*, vol. 25, no. 23, p. 7294, 2025.
- [25] Z. Wang, Z. Zhang, W. Wu, Z. Zhang, X. Xu, and H. Li, "State analysis of grouped smart meters driven by interpretable random forest," *Electronics*, vol. 14, no. 15, 2025.
- [26] Q.-A. Wang, J. Chen, Y. Ni, Y. Xiao, N. Liu, S. Liu, and W. Feng, "Application of bayesian networks in reliability assessment: A systematic literature review," *Structures*, vol. 71, p. 108098, 2025.
- [27] A. Ben Mrad, V. Delcroix, S. Piechowiak, and P. Leicester, "From Information to Evidence in a Bayesian Network," in *Probabilistic Graphical Models* (L. C. van der Gaag and A. J. Feelders, eds.), (Cham), pp. 33–48, Springer International Publishing, 2014.
- [28] C. Chan, "A proportional hazards approach to correlate SiO₂/sub 2-breakdown voltage and time distributions," *IEEE Transactions on Reliability*, vol. 39, pp. 147–150, June 1990.
- [29] D. Y. Lin, "On the Breslow estimator," *Lifetime Data Analysis*, vol. 13, pp. 471–480, Dec. 2007.
- [30] D. Y. LIN, L. J. WEI, and Z. YING, "Checking the cox model with cumulative sums of martingale-based residuals," *Biometrika*, vol. 80, pp. 557–572, 09 1993.
- [31] M. T. Todinov, *Risk-based reliability analysis and generic principles for risk reduction*. Elsevier, 2006.
- [32] Y. Li, X. Xiao, Z. Zhang, and W. Liu, "A feature engineering framework for smart meter group failure rate prediction," *Mathematics*, vol. 13, no. 15, p. 2472, 2025.
- [33] R. Sharma, J. S. Matharu, and K. S. Parmar, "A survey on particle swarm optimization: Evolution, adaptations and practical implementations," *Applied Soft Computing*, p. 114016, 2025.
- [34] A. Ly, M. Marsman, J. Verhagen, R. P. Grasman, and E.-J. Wagenmakers, "A tutorial on fisher information," *Journal of Mathematical Psychology*, vol. 80, pp. 40–55, 2017.
- [35] M. Borrotti, "Quantifying uncertainty with conformal prediction for heating and cooling load forecasting in building performance simulation," *Energies*, vol. 17, no. 17, p. 4348, 2024.
- [36] R. Moghaddass and J. Wang, "A hierarchical framework for smart grid anomaly detection using large-scale smart meter data," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5820–5830, 2017.
- [37] J. M. Zhang, M. Harman, L. Ma, and Y. Liu, "Machine learning testing: Survey, landscapes and horizons," *IEEE Transactions on Software Engineering*, vol. 48, no. 1, pp. 1–36, 2020.
- [38] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (methodological)*, vol. 34, no. 2, pp. 187–202, 1972.
- [39] Q. Chen, H. Wu, L. B. Ware, and T. Koyama, "A Bayesian Approach for the Cox Proportional Hazards Model with Covariates Subject to Detection Limit," *International journal of statistics in medical research*, vol. 3, pp. 32–43, Jan. 2014.
- [40] J. Krausangka and M. J. Druzdzel, "A Bayesian network interpretation of the Cox's proportional hazard model," *International Journal of Approximate Reasoning*, vol. 103, pp. 195–211, Dec. 2018.
- [41] F. Faghiri and A. Kohansal, "Cox proportional hazards model with Bayesian neural network for survival prediction," *Scientific Reports*, vol. 15, p. 31581, Aug. 2025.
- [42] P. J. Heagerty and Y. Zheng, "Survival model predictive accuracy and roc curves," *Biometrics*, vol. 61, no. 1, pp. 92–105, 2005.
- [43] H. M. Attaullah, M. Harris, I. U. Khan, M. M. Alam, and M. M. Su'ud, "Comparative analysis of ML-based intrusion detection system for cyber-physical UAV system," *Journal of Computational and Cognitive Engineering*, vol. 5, no. 1, pp. 44–52, 2026.
- [44] X. Chen, Y. Lei, Y.-F. Li, S. Parkinson, X. Li, J. Liu, F. Lu, H. Wang, Z. Wang, B. Yang, S. Ye, and Z. Zhao, "Large models for machine monitoring and fault diagnostics: Opportunities, challenges, and future direction," *Journal of Dynamics, Monitoring and Diagnostics*, vol. 4, no. 2, pp. 76–90, 2025.