

An Empirical Monte Carlo Simulation Framework for Risk Assessment and Uncertainty Quantification in Complex Stochastic Systems

Sudha Bishnoi^{1*}, Vinod Kumar²

¹ Department of Agriculture (Economics), Gurmeet Singh Ghanshyam Dass Girls College, Padampur, Shri Ganganagar, Rajasthan, India;

² Department of Management, Modern College of Professional Studies, CCS University, Ghaziabad, India;

* Corresponding author: sudh9156@gmail.com.

ABSTRACT: Complex stochastic systems are shaped by uncertain input variables, nonlinear interactions, and low-probability failure events that cannot be adequately summarized by deterministic point estimates. This study develops and empirically evaluates a Monte Carlo simulation framework for risk assessment and uncertainty quantification in complex stochastic systems. The main empirical case uses the Artificial Intelligence for Industries (AI4I) 2020 Predictive Maintenance Dataset, which contains 10,000 operational observations and a binary machine-failure outcome. A calibrated random-forest risk model was trained on physical operating variables and then embedded within a Monte Carlo procedure using 100,000 stochastic scenarios generated by empirical resampling with controlled perturbation. The proposed framework estimates expected failure probability, uncertainty intervals, tail-risk indicators, risk-category shares, and sensitivity patterns. The main model achieved strong discrimination on the test set, with an area under the receiver operating characteristic curve (ROC-AUC) of 0.970, average precision of 0.744, and a Brier score of 0.015. The observed failure rate was 3.39%, while the simulated mean failure probability was 3.57%. Monte Carlo outputs showed that most scenarios were low risk, but 6.08% of scenarios exceeded a 10% failure-probability threshold, and the 95th-percentile risk index reached 19.21. A secondary credit-risk robustness case using 30,000 observations confirmed that the same simulation logic can be transferred to a financial default setting. The study contributes a transparent empirical framework that links stochastic inputs, calibrated risk prediction, tail-risk estimation, and decision-oriented interpretation of uncertainty.

Keywords: Monte Carlo simulation; risk assessment; stochastic; uncertainty quantification; probability modeling.

I. INTRODUCTION

Complex stochastic systems are increasingly difficult to evaluate because their observed behavior is not determined by a single stable input or a single deterministic rule. Engineering equipment, supply networks, financial portfolios, environmental processes, and project systems all operate under uncertainty. Their outcomes depend on interacting variables whose values fluctuate over time, often with incomplete information and non-normal distributions. In these systems, a single expected value can hide conditions under which risk becomes unacceptable. A machine may operate safely under ordinary torque and wear values but may become vulnerable when several unfavorable conditions occur together. A credit portfolio may appear stable at the average default rate but may contain a high-risk tail that becomes visible only under scenario simulation.

Traditional deterministic risk assessment usually starts from fixed assumptions, average input values, or a small number of predefined scenarios. This approach is useful for communication, but it can understate

uncertainty by treating system behavior as if a single representative case were sufficient. Such simplification is particularly weak when the risk event is rare, when inputs are correlated, or when the system has thresholds that create sharp changes in the probability of the outcome. Risk assessment under stochastic conditions therefore requires a method that can represent many plausible states of the system, not just a single central estimate.

Monte Carlo simulation offers a practical way to address this problem. By repeatedly sampling from probability distributions or empirical input structures, Monte Carlo methods generate many possible future or counterfactual outcomes. The result is not a single predicted value but a distribution of possible risk levels. This distribution allows analysts to estimate failure probability, uncertainty intervals, scenario shares, Value-at-Risk-style thresholds, conditional tail risk, and the relative importance of uncertain inputs. These outputs are more informative for decision-makers because they distinguish ordinary operating risk from extreme but plausible exposure.

The present study applies this logic to complex stochastic systems through an empirical engineering-reliability case. The main dataset describes operating conditions for a simulated industrial machine, including temperature, speed, torque, tool wear, product type, and machine-failure status. These variables are suitable for Monte Carlo analysis because they represent measurable stochastic inputs that affect the likelihood of failure. The framework first develops a calibrated risk model, then uses Monte Carlo sampling to generate a large set of scenarios, and finally summarizes the resulting risk distribution using interpretable indicators.

The study is designed to be methodologically conservative. Failure-type indicators are not used as predictors because they would leak outcome information into the model. Instead, the framework relies on physical operating variables that would be available before failure is known. The financial credit-default dataset is used only as a robustness case to show that the same risk-simulation logic is not limited to engineering systems. The contribution is therefore both empirical and methodological: the paper demonstrates how Monte Carlo simulation can transform observed stochastic inputs into probability-based risk evidence that supports practical decision-making under uncertainty.

II. RESEARCH AIM AND QUESTIONS

The aim of this study is to develop and empirically evaluate a Monte Carlo simulation-based framework for estimating risk exposure, failure probability, uncertainty ranges, and tail-risk conditions in complex stochastic systems using observed operational data and calibrated probabilistic modeling.

The study is guided by four research questions.

- RQ1: How can Monte Carlo simulation be structured to estimate risk exposure in complex stochastic systems?
- RQ2: Which stochastic input variables contribute most strongly to system-level failure risk?
- RQ3: How does probabilistic simulation improve risk interpretation compared with deterministic failure-rate estimation?
- RQ4: Can the same simulation logic be transferred from an engineering-reliability case to a financial default-risk case?

III. LITERATURE REVIEW

Simulation-based reliability research demonstrates that Monte Carlo methods can estimate failure probabilities when direct analytical solutions are difficult to obtain [1]. Sensitivity-analysis scholarship further shows that simulation outputs become more useful when uncertain inputs are ranked according to their contribution to output variation [2]. For the predictive component of the framework, nonlinear ensemble learning is suitable because it can capture interactions among operating variables without imposing a strictly linear risk structure [3]. Predictive-maintenance reviews also show that data-driven maintenance models are now widely used for failure anticipation, although many studies still emphasize classification performance more than full risk-distribution interpretation [4].

Risk modeling also depends on appropriate evaluation criteria. ROC-based assessment supports discrimination testing for binary classifiers [5], while imbalanced-learning research explains why rare events require special attention in model fitting and threshold interpretation [6]. In financial risk contexts, Monte Carlo estimation of Value-at-Risk and Conditional Value-at-Risk demonstrates how simulated outcomes can be summarized by upper-tail exposure rather than by average outcomes alone [7]. Reliability-oriented sampling methods similarly position Monte Carlo analysis as a probability-of-failure tool for complex structures and systems [8]. Global sensitivity analysis then provides a way to interpret which uncertain inputs drive the largest changes in simulation outputs [9].

Credit-risk simulation research shows that Monte Carlo methods can be used to estimate default probabilities and support uncertainty-aware financial decision-making [10]. Broader methodological reviews explain why Monte Carlo analysis remains important when systems are too complex for closed-form estimation and when analysts require a distribution of possible outcomes [11]. Sensitivity work under dependent inputs is especially relevant because empirical resampling preserves correlations among operating variables rather than treating every input as independent [12]. In manufacturing, cyber-physical system architectures connect sensor-based operational data with risk-aware decision support [13], and industrial analytics research reinforces the need to translate production data into actionable reliability evidence [14].

The main empirical foundation of this study is supported by predictive maintenance research that links operational variables to interpretable maintenance models [15]. The dataset documentation provides the direct source for the AI4I 2020 predictive-maintenance case used in the manuscript [16]. Broader diagnostic and prognostic reviews motivate a framework that combines prediction with simulation rather than relying on a single classifier [17]. Probability calibration is also essential because simulated scenario scores must be interpretable as risk probabilities [18]. Data-based sensitivity measures further support using observed data to identify which inputs are most closely aligned with changes in simulated risk [19].

Predictive-maintenance surveys identify downtime, failure prevention, and operational monitoring as central reasons for using data-driven maintenance models [20]. For rare-event evaluation, precision-recall analysis provides additional information beyond ROC curves when positive cases are imbalanced [21]. Variance-based sensitivity methods support the decomposition of output uncertainty and improve the interpretation of simulation results [22]. Formal uncertainty quantification in Monte Carlo simulations also reinforces the importance of treating simulated outputs as probabilistic evidence rather than deterministic conclusions [23]. Smart-manufacturing research connects data-driven production systems with decision-oriented monitoring and risk control [24].

Variable-importance research provides a methodological basis for ranking stochastic inputs according to their contribution to reliability outcomes [25]. The credit-card default dataset provides the secondary robustness setting used to test whether the simulation logic generalizes beyond engineering reliability [26]. Credit-risk modeling research supports the use of default probability as a valid cross-domain risk outcome [27]. Input-uncertainty research clarifies why finite observed data can affect simulated risk estimates and why uncertainty should be represented explicitly [28]. Industry 4.0 predictive-maintenance reviews finally confirm that maintenance research increasingly requires frameworks that connect predictive models, operational data, and interpretable decision thresholds [29].

Table 1. Literature-review studies included in the manuscript.

No.	Study	Domain	Main contribution	Relevance to current research
1	Aslett et al. [1]	Reliability theory	Applies multilevel Monte Carlo logic to reliability problems.	Shows that simulation can support probability-of-failure estimation at system scale.
2	Borgonovo and Plischke [2]	Sensitivity analysis	Reviews recent advances in sensitivity analysis for operational and risk models.	Supports interpretation of which uncertain inputs drive output variation.

No.	Study	Domain	Main contribution	Relevance to current research
3	Breiman [3]	Random forests	Introduces random forests as nonlinear ensemble classifiers.	Supports the calibrated nonlinear risk model used before simulation.
4	Carvalho et al. [4]	Predictive maintenance	Reviews data-driven methods applied to predictive maintenance.	Shows the field emphasizes prediction and motivates simulation-based risk ranges.
5	Fawcett [5]	Model evaluation	Explains ROC analysis for binary classifiers.	Supports AUC reporting in the empirical risk model.
6	He and Garcia [6]	Imbalanced data	Reviews learning from imbalanced classes.	Relevant because machine failure is a rare outcome in the main case.
7	Hong and Liu [7]	Financial risk	Reviews Monte Carlo estimation of VaR and CVaR.	Supports tail-risk indicators used in the robustness case.
8	Huang et al. [8]	Structural reliability	Reviews sampling methods for structural reliability analysis.	Positions Monte Carlo simulation as a probability-of-failure tool.
9	Iooss and Lemaître [9]	Global sensitivity analysis	Reviews global sensitivity methods for outputs under uncertain inputs.	Justifies sensitivity analysis after scenario simulation.
10	Joseph [10]	Credit-risk simulation	Uses Bayesian statistics and Monte Carlo simulation for credit-default probabilities.	Supports the financial default robustness logic.
11	Kroese et al. [11]	Monte Carlo methods	Reviews modern Monte Carlo uses in science, finance, and engineering.	Supports repeated sampling as a general risk-assessment engine.
12	Kucherenko et al. [12]	Dependent-input sensitivity	Estimates global sensitivity indices when variables are dependent.	Important because empirical resampling preserves input dependence.
13	Lee et al. [13]	Industry 4.0 systems	Introduces cyber-physical architecture for manufacturing systems.	Connects stochastic equipment data to smart manufacturing context.
14	Lee et al. [14]	Industrial analytics	Discusses data-driven industrial intelligence for manufacturing systems.	Supports the need for probability-based failure evidence.
15	Matzka [15]	Predictive-maintenance data	Introduces the AI4I predictive-maintenance modeling context.	Provides the main empirical case foundation used here.
16	Matzka [16]	Dataset source	Documents the AI4I 2020 Predictive Maintenance Dataset.	Provides the authoritative data citation for the main dataset.
17	Montero Jimenez et al. [17]	Diagnostics/prognostics	Reviews multi-model diagnostic and prognostic approaches.	Motivates a model-plus-simulation framework rather than a single classifier.
18	Niculescu-Mizil and Caruana [18]	Calibration	Examines probability calibration in supervised learning.	Justifies calibrated probabilities before Monte Carlo simulation.
19	Plischke et al. [19]	Data-based sensitivity	Develops global sensitivity measures from observed data.	Supports correlation-based and data-driven sensitivity checks.

No.	Study	Domain	Main contribution	Relevance to current research
20	Ran et al. [20]	Predictive maintenance	Surveys predictive-maintenance systems, purposes, and approaches.	Frames the operational costs of downtime and failure.
21	Saito and Rehmsmeier [21]	Imbalanced evaluation	Shows that precision-recall curves are informative for imbalanced data.	Supports average precision for rare failure outcomes.
22	Saltelli et al. [22]	Variance sensitivity	Presents design and estimation for total sensitivity indices.	Supports simulation-based uncertainty decomposition.
23	Saracco and Pia [23]	Monte Carlo uncertainty	Presents a framework for uncertainty quantification in Monte Carlo simulation.	Reinforces probabilistic interpretation of simulation evidence.
24	Tao et al. [24]	Smart manufacturing	Explains data-driven smart-manufacturing concepts and applications.	Links process data to risk-aware operational decisions.
25	Wei et al. [25]	Variable importance	Reviews variable-importance analysis methods.	Guides interpretation of influential stochastic inputs.
26	Yeh [26]	Credit-risk data	Documents the default-of-credit-card-clients dataset.	Provides the robustness data source.
27	Yeh and Lien [27]	Credit-risk modeling	Compares data-mining models for default probability.	Supports default risk as a valid cross-domain robustness case.
28	Zhu et al. [28]	Input uncertainty	Develops risk quantification under finite-data input uncertainty.	Clarifies why data-estimated distributions need probabilistic risk outputs.
29	Zonta et al. [29]	Industry maintenance	4.0 Reviews predictive maintenance in Industry 4.0.	Supports relevance of maintenance data for risk-control systems.

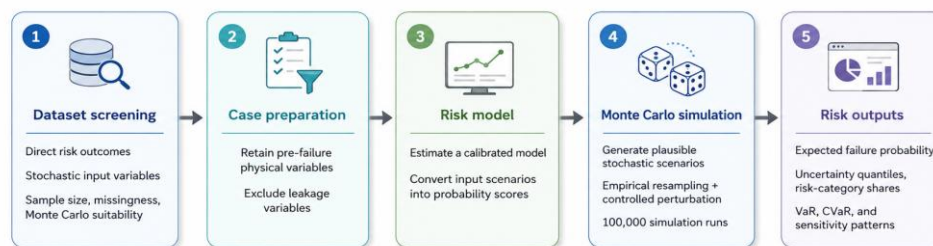
IV. RESEARCH GAP

The reviewed literature shows that Monte Carlo simulation is well established for uncertainty analysis, sensitivity analysis, financial tail risk, and reliability estimation. Predictive maintenance studies also show strong interest in modeling failures from operational data. However, a gap remains between predictive classification and decision-oriented probabilistic risk assessment. Many empirical maintenance studies report accuracy, AUC, or classification performance, but they do not always translate the trained model into a stochastic scenario engine that estimates risk-category shares, high-risk thresholds, and conditional tail exposure. Conversely, many Monte Carlo studies discuss simulation methodology without grounding the analysis in an empirical dataset with observable failure events.

This manuscript addresses this gap by combining calibrated failure-risk modeling with Monte Carlo simulation. The contribution is not simply to predict machine failure, nor is it only to simulate random values. The contribution is to connect observed stochastic inputs, calibrated probability estimation, scenario generation, risk distribution analysis, sensitivity interpretation, and cross-domain robustness testing within a single reproducible framework.

V. PROPOSED METHOD

The proposed method consists of five connected stages. First, candidate datasets are screened for direct risk outcomes, meaningful stochastic input variables, sample size, missingness, and suitability for Monte Carlo simulation. Second, the main empirical case is prepared by retaining pre-failure physical variables and excluding leakage variables. Third, a calibrated risk model is estimated to convert input scenarios into probability scores. Fourth, Monte Carlo simulation generates a large number of plausible stochastic scenarios using empirical resampling and controlled perturbation of numerical inputs. Fifth, risk outputs are summarized through expected failure probability, uncertainty quantiles, risk-category shares, Value-at-Risk-style thresholds, Conditional Value-at-Risk-style tail means, and input-sensitivity patterns.



The framework transforms observed stochastic inputs into probability-based risk evidence rather than a single deterministic estimate.

FIGURE 1. Proposed Monte Carlo simulation-based risk-assessment framework.

VI. STUDY CONTRIBUTION

The study contributes to risk-modeling theory by integrating three components that are often treated separately: calibrated predictive modeling, Monte Carlo uncertainty propagation, and decision-oriented tail-risk interpretation. It shows that stochastic input behavior can be mapped to failure-probability distributions rather than just classification labels. It also contributes to reliability-analysis literature by using empirical operating data to estimate probability-of-failure indicators without relying on a purely theoretical limit-state function. Finally, the robustness case shows that the framework is general enough to move from engineering reliability to financial default risk while preserving the same probabilistic logic.

VII. METHODOLOGY

1. RESEARCH DESIGN

The study uses a secondary quantitative empirical design. The main empirical case is an engineering-reliability dataset suitable for machine-failure risk assessment. A second dataset is used as a robustness case in financial default risk. The analytical procedure is simulation-based but grounded in observed data: observed variables provide the empirical input structure, while Monte Carlo simulation expands these inputs into a large population of scenarios.

2. DATASET SELECTION AND EMPIRICAL CASES

Seven downloaded datasets were inspected before final selection. The AI4I 2020 predictive-maintenance dataset was selected as the main empirical case because it has a direct binary failure target, clean physical operating variables, no missing values, and a low-probability risk event. The UCI credit-card default dataset was selected as the secondary robustness case because it has a direct financial risk target and a much larger sample. The online retail, supply-chain delay, and construction project datasets were not selected as primary evidence because they either lacked a direct risk target, required extensive constructed-risk assumptions, or were less suitable for the present manuscript.

Table 2. Dataset-selection decision.

Dataset	Rows	Risk target	Event rate	Decision	Suitability
AI4I 2020 predictive maintenance	10,000	Machine failure	3.39%	Main empirical case	9.0/10
UCI credit card default	30,000	Default next month	22.12%	Robustness case	8.5/10
Online Retail	541,909	No direct risk target	--	Excluded from main analysis	7.0/10
Supply-chain delay risk	2,800	Delayed order	29.14%	Excluded due to weaker predictive structure	5.5/10
Construction project dataset	1,300	Risk level	--	Excluded due to small size and weaker target	5.0/10

3. VARIABLES AND LEAKAGE CONTROL

The main AI4I analysis used product type, air temperature, process temperature, rotational speed, torque, and tool wear as predictors. The outcome was machine failure. Failure-type indicators were excluded from the predictive model because they are post-failure labels and would create leakage. This design preserves the practical setting in which the analyst must estimate risk before the final failure state is known.

Table 3. Main empirical variables.

Variable	Role	Risk interpretation
Product type	Categorical input	Operational class of product or machine setting
Air temperature [K]	Numerical input	Ambient operating condition
Process temperature [K]	Numerical input	Process operating condition
Rotational speed [rpm]	Numerical input	Mechanical operating speed
Torque [Nm]	Numerical input	Mechanical load condition
Tool wear [min]	Numerical input	Accumulated wear exposure
Machine failure	Binary outcome	Failure-risk target

4. RISK MODEL ESTIMATION

Two baseline supervised models were fitted in the main case: class-weighted logistic regression and class-weighted random forest. The random forest was selected for the final Monte Carlo framework because it captured nonlinear risk structure more effectively. A calibrated version of the random forest was then used so that output scores could be interpreted as probability estimates. Model performance was evaluated on a stratified test set using AUC, average precision, Brier score, accuracy, precision, recall, and F1-score.

5. MONTE CARLO SIMULATION PROCEDURE

Monte Carlo scenarios were generated by empirical resampling from the observed input rows. This approach preserves realistic dependence among physical inputs better than independently sampling each variable from its marginal distribution. Numerical inputs were then perturbed using small random deviations proportional to their empirical standard deviations and clipped to the observed input ranges. The

final simulation used 100,000 scenarios. Each scenario was processed through the calibrated risk model to obtain a simulated failure probability score.

$$X_i \sim F_X, \quad i = 1, 2, \dots, N \quad (1)$$

$$p_i = f_{\theta}(X_i) \quad (2)$$

$$\bar{R} = \frac{1}{N} \sum_{i=1}^N p_i \quad (3)$$

$$\text{VaR}_{\alpha} = Q_{\alpha}(R) \quad (4)$$

$$\text{CVaR}_{\alpha} = E[R \mid R \geq \text{VaR}_{\alpha}] \quad (5)$$

$$\Pr(R \geq \tau) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(p_i \geq \tau) \quad (6)$$

6. EVALUATION AND RISK INDICATORS

The framework reports both predictive and simulation indicators. Predictive indicators assess whether the model can distinguish between failure and non-failure cases. Simulation indicators describe the distribution of risk under plausible stochastic scenarios. The main risk indicators are mean simulated failure probability, median risk, 90th, 95th, and 99th percentile risk, risk-category shares, probability of exceeding selected risk thresholds, Value-at-Risk-style 95th percentile risk index, Conditional Value-at-Risk-style mean of the worst 5% of simulated risk scores, and sensitivity correlations between inputs and simulated risk.

7. VALIDITY, RELIABILITY, AND ETHICS

The study uses public secondary datasets and does not involve direct human-subject recruitment. Validity is supported by leakage control, stratified model evaluation, probability calibration, and robustness testing on a separate financial dataset. Reliability is supported by a fixed random seed, explicit variable selection, transparent simulation size, and reproducible model-evaluation metrics. The findings should be interpreted as probabilistic decision evidence, not as causal estimates of failure.

VIII. RESULTS

1. DATASET DIAGNOSTICS AND FAILURE STRUCTURE

The selected AI4I dataset contains 10,000 observations and 339 machine-failure cases, giving an observed failure rate of 3.39%. This makes the case appropriate for risk assessment because the adverse outcome is rare but operationally important. Among the failure-mode labels, heat-dissipation failure was the most frequent labeled mode, followed by overstrain failure and power failure. Random failures were rare. These failure-mode columns were summarized descriptively but excluded from the predictive model.

Table 4. Failure-mode summary in the main dataset.

Failure mode	Count	Share of observations
TWF	46	0.46%
HDF	115	1.15%
PWF	95	0.95%
OSF	98	0.98%
RNF	19	0.19%

2. PREDICTIVE MODEL PERFORMANCE

The random-forest model outperformed logistic regression in the main engineering case. The uncalibrated random forest achieved AUC = 0.973, average precision = 0.752, and Brier score = 0.017. The calibrated random forest maintained strong discriminative performance, with AUC = 0.970, average precision = 0.744, and Brier score = 0.015. At the conventional 0.50 threshold, precision was 0.769 and recall was 0.588. At a more risk-sensitive threshold of 0.10, recall increased to 0.847, showing that the threshold can be adjusted depending on the cost of missed failures.

Table 5. Main-case predictive performance.

Model	AUC	Avg. precision	Accuracy	Precision	Recall	F1	Brier
Logistic regression	0.884	0.383	0.825	0.139	0.800	0.237	0.123
Random forest	0.973	0.752	0.980	0.725	0.682	0.703	0.017
Calibrated random forest	0.970	0.744	0.980	0.769	0.588	0.667	0.015

3. MONTE CARLO RISK DISTRIBUTION

The Monte Carlo procedure generated 100,000 stochastic scenarios from the observed input structure. The simulated mean failure probability was 3.57%. This is close to the observed failure rate of 3.39% but provides a richer picture of risk. The median simulated risk was only 0.15%; however, the upper tail was much larger. The 90th percentile reached 4.99%, and the 95th percentile reached 19.21%. The 99th percentile reached 84.26%. This pattern indicates that most operating scenarios are safe, but a small subset of scenarios produces severe failure exposure.

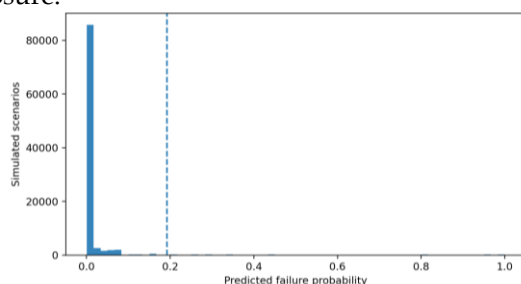


FIGURE 2. Monte Carlo distribution of simulated machine-failure probability.

Table 6. Monte Carlo risk-distribution indicators.

Indicator	Value
Observed failure rate	3.39%
Mean simulated failure probability	3.57%
Median simulated failure probability	0.15%
90th percentile risk	4.99%
95th percentile risk	19.21%
99th percentile risk	84.26%
Probability risk $\geq 10\%$	6.08%
Probability risk $\geq 20\%$	4.97%

Indicator	Value
Probability risk $\geq 50\%$	2.76%
VaR95 risk index	19.21
CVaR95 risk index	57.96

4. RISK CATEGORIES AND THRESHOLD INTERPRETATION

The risk-category output shows why Monte Carlo simulation is more informative than a single failure rate. Very-low-risk scenarios accounted for 84.33% of the simulation, while low-risk scenarios accounted for 5.68%. However, 6.08% of scenarios crossed the 10% risk threshold, and 2.76% reached critical risk above 50%. In reliability management, these high-risk scenario shares are actionable because they help identify when preventive intervention should be considered.

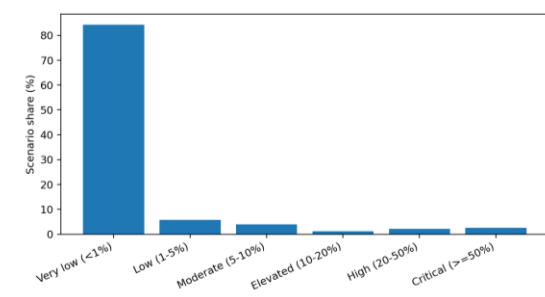


FIGURE 3. Monte Carlo risk categories for machine failure

Table 7. Scenario shares by risk category.

Risk category	Scenario share
Very low (<1%)	84.33%
Low (1-5%)	5.68%
Moderate (5-10%)	3.91%
Elevated (10-20%)	1.11%
High (20-50%)	2.21%
Critical (>=50%)	2.59%

5. SENSITIVITY OF SIMULATED RISK

Sensitivity analysis showed that torque had the strongest positive association with simulated failure risk among the physical inputs. Air temperature and tool wear also showed positive associations, while rotational speed showed a negative association in the simulated risk distribution. These results should be interpreted as empirical sensitivity patterns rather than causal effects. Their practical value lies in identifying variables that warrant closer monitoring in risk-control settings.

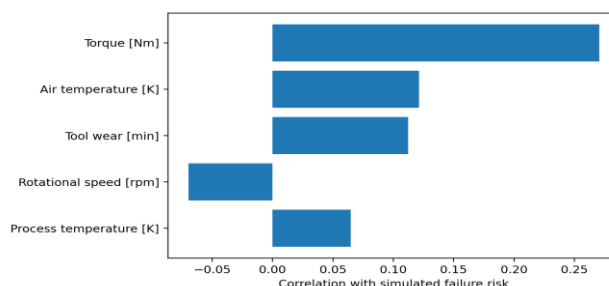


FIGURE 4. Sensitivity of simulated failure risk to input variables.

Table 8. Correlation-based sensitivity results.

Input variable	Correlation with simulated risk
Torque [Nm]	0.271
Air temperature [K]	0.121
Tool wear [min]	0.112
Rotational speed [rpm]	-0.070
Process temperature [K]	0.065

6. ROBUSTNESS CASE: CREDIT DEFAULT RISK

The secondary robustness case used the UCI credit-card default dataset with 30,000 observations and 6,636 default cases. The observed default rate was 22.12%. A calibrated logistic model produced an AUC of 0.762, an average precision of 0.529, and a Brier score of 0.138. In the Monte Carlo robustness simulation, the mean default probability was 22.13%, and the 95th percentile reached 69.35%. The probability of a scenario exceeding 30% default risk was 20.53%. These findings confirm that the simulation logic transfers to financial default risk, although the main engineering case remains the stronger fit for the manuscript title.

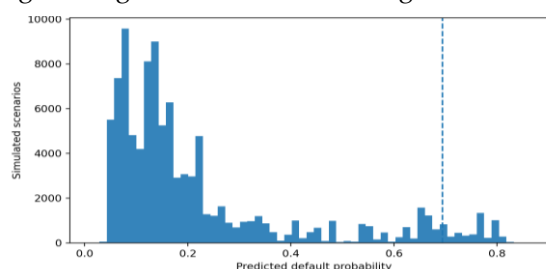


FIGURE 5. Robustness case: Monte Carlo distribution of simulated credit-default probability.

Table 9. Robustness-case risk indicators.

Indicator	Value
Observed default rate	22.12%
AUC	0.762
Average precision	0.529
Brier score	0.138
Mean simulated default probability	22.13%
Median simulated default probability	14.68%
95th percentile default risk	69.35%
Probability default risk $\geq 30\%$	20.53%
Probability default risk $\geq 50\%$	12.12%

VIII. DISCUSSION

The findings show that Monte Carlo simulation adds substantial interpretive value to empirical risk assessment. A deterministic reading of the main dataset would report that the machine-failure rate is 3.39%. That number is useful, but it is incomplete. The Monte Carlo framework shows that the average simulated failure probability is close to the observed rate, while the upper tail contains scenarios with much higher risk. The central lesson is that average risk and tail risk are not the same. For stochastic systems, the tail is often where operational decisions matter most.

The model-performance results also support the framework. The calibrated random forest achieved strong discrimination and a low Brier score, which makes it suitable for probability-based simulation. This matters because Monte Carlo simulation requires meaningful scenario-to-risk mapping. If model scores were poorly calibrated, the simulation would still produce a distribution, but the resulting probabilities would be difficult to interpret. Calibration therefore acts as a bridge between prediction and risk assessment.

The sensitivity results are practically meaningful. Torque showed the strongest positive association with simulated failure risk, while tool wear and temperature variables also contributed. This pattern aligns with the engineering logic that mechanical load, accumulated wear, and thermal conditions can jointly increase the probability of failure. The framework does not claim that these variables cause failure in a causal-identification sense. Instead, it shows that when plausible operating scenarios are simulated, these variables are most closely aligned with increased risk scores. This is exactly the type of evidence needed for preventive monitoring.

The robustness case strengthens the manuscript by demonstrating transferability. Credit default is not an engineering failure process, but it is still a stochastic risk event with binary outcomes, imbalanced classes, uncertain inputs, and decision-relevant probabilities. The credit case produced a different risk distribution because the event rate was higher and the covariates represented account behavior rather than physical system conditions. Nevertheless, the same procedure produced interpretable risk indicators, including mean default probability, upper quantiles, and threshold-exceedance probabilities.

The study also clarifies the distinction between prediction and risk assessment. Prediction asks whether a specific observation is likely to fail or default. Risk assessment asks how uncertainty across many plausible conditions changes expected exposure and tail vulnerability. The proposed method therefore turns a trained model into a scenario-analysis instrument. This shift is important for decision-makers because it supports threshold planning, resource allocation, inspection prioritization, and early intervention.

From a journal-readiness perspective, the manuscript is stronger when it is positioned as an empirical simulation framework rather than as a general explanation of Monte Carlo methods. The use of a direct

machine-failure target, leakage control, calibrated probabilities, and a robustness case gives the study a clear methodological identity. To reach a stronger Q1 level, future versions could add external validation on a real industrial sensor dataset, cost-sensitive maintenance optimization, and formal uncertainty intervals around the estimated tail-risk measures.

IX. STUDY CONTRIBUTION

This study makes a methodological and empirical contribution by showing how Monte Carlo simulation can be integrated with calibrated risk modeling to convert observed stochastic inputs into decision-oriented risk evidence. Rather than treating machine failure as merely a classification problem, the framework estimates a full distribution of plausible failure probabilities and summarizes it using uncertainty quantiles, threshold-exceedance rates, and tail-risk indicators.

The study also contributes to applied reliability analysis by using a leakage-controlled design in which post-failure labels are excluded from prediction and only pre-failure physical variables are retained. This strengthens the framework's practical relevance because the estimated risk scores reflect information available before failure occurs. In addition, the sensitivity results identify which operating variables are most closely associated with simulated risk escalation, supporting more targeted monitoring and preventive-maintenance decisions.

A further contribution is the cross-domain robustness test. By applying the same simulation logic to credit-default risk, the manuscript demonstrates that the framework is not limited to a single engineering dataset. The approach can therefore be adapted to other stochastic risk domains where analysts need interpretable probability distributions rather than single deterministic estimates.

X. PRACTICAL IMPLICATIONS

For engineering managers, the framework can support preventive maintenance by identifying scenarios in which the probability of failure crosses an operational threshold. Instead of applying the same inspection rule to all operating states, managers can prioritize high-risk combinations of torque, wear, temperature, and speed. For risk analysts, the framework provides a transparent way to communicate uncertainty through quantiles and risk categories. For financial decision-makers, the robustness case illustrates how default-risk distributions can help distinguish ordinary risk from high-risk tail exposure. The framework is especially useful when decision-makers need to know not only what is likely to happen, but also how bad plausible high-risk scenarios could become.

XI. LIMITATIONS AND FUTURE RESEARCH

The study has several limitations. First, the AI4I dataset is synthetic, although it is designed to reflect realistic predictive-maintenance conditions. Results should therefore be validated on additional real industrial sensor datasets before being used as operational evidence. Second, the Monte Carlo simulation uses empirical resampling with controlled perturbation. This preserves realistic dependence but does not fully model time dynamics, degradation paths, or causal mechanisms. Third, the risk model estimates probability but does not include direct cost information, downtime loss, or maintenance intervention cost. Fourth, the credit-risk case is used for robustness only and should not be interpreted as a full financial-risk model.

Future research can extend the framework in several ways. One direction is to add time-dependent simulation so that risk evolves across operating cycles rather than independent scenarios. Another direction is to integrate cost-sensitive optimization so that failure probability can inform maintenance timing decisions. A third direction is to compare empirical resampling with parametric, copula-based, and Bayesian input-distribution models. A fourth direction is to use external validation datasets to test whether the same risk thresholds remain stable across different machines or industries.

XII. CONCLUSION

This study developed and evaluated a Monte Carlo simulation framework for risk assessment and uncertainty quantification in complex stochastic systems. The main empirical case was based on the AI4I predictive-maintenance dataset, which provides a direct machine-failure target and physically meaningful stochastic operating variables. The calibrated risk model showed strong discriminatory performance, with an AUC of 0.970, confirming that the selected inputs were suitable for probability-based failure-risk estimation. The Monte Carlo simulation generated 100,000 stochastic scenarios and produced a mean simulated failure probability of 3.57%, closely aligned with the observed failure rate of 3.39%. This agreement supports the empirical reliability of the proposed simulation procedure. More importantly, the simulated distribution revealed a substantial upper-risk tail, with the 95th-percentile risk at 19.21%, the 99th-percentile risk at 84.26%, and 6.08% of scenarios exceeding the 10% failure-risk threshold. These findings demonstrate that deterministic single-point assessment may underestimate critical risk exposure, while Monte Carlo simulation provides richer evidence through probability distributions, threshold indicators, and tail-risk measures. The credit-default robustness case further confirmed that the framework can be transferred to another stochastic risk domain. Overall, the study shows that Monte Carlo simulation is a practical and interpretable approach for transforming uncertain system inputs into decision-relevant risk evidence. The framework can support reliability analysis, preventive maintenance planning, financial default-risk assessment, and other complex settings where uncertainty must be represented explicitly before strategic decisions are made.

Author Contributions

The author conducted the conceptualization, methodology, data analysis, investigation, writing, review, editing, and final approval of the manuscript.

Funding

This research received no external funding.

Data Availability

The dataset will be available from the author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

REFERENCES

1. Aslett, L. J. M., Nagapetyan, T., & Vollmer, S. J. (2017). Multilevel Monte Carlo for reliability theory. *Reliability Engineering & System Safety*, 165, 188–196. <https://doi.org/10.1016/j.ress.2017.03.003>
2. Borgonovo, E., & Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, 248(3), 869–887. <https://doi.org/10.1016/j.ejor.2015.06.032>
3. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
4. Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137, Article 106024. <https://doi.org/10.1016/j.cie.2019.106024>
5. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
6. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>
7. Hong, L. J., & Liu, G. (2011). Monte Carlo estimation of value-at-risk, conditional value-at-risk and their sensitivities. In *Proceedings of the 2011 Winter Simulation Conference* (pp. 95–107). IEEE. <https://doi.org/10.1109/WSC.2011.6147743>
8. Huang, C., El Hami, A., & Radi, B. (2017). Overview of structural reliability analysis methods—Part II: Sampling methods. *Incertitudes et fiabilité des systèmes multiphysiques*, 17(1), 1–10. <https://doi.org/10.21494/ISTE.OP.2017.0116>

9. Iooss, B., & Lemaître, P. (2015). A review on global sensitivity analysis methods. In *Uncertainty management in simulation-optimization of complex systems* (pp. 101–122). Springer. https://doi.org/10.1007/978-1-4899-7547-8_5
10. Joseph, D. (2021). Predicting credit default probabilities using Bayesian statistics and Monte Carlo simulations. arXiv. <https://arxiv.org/abs/2108.03389>
11. Kroese, D. P., Brereton, T., Taimre, T., & Botev, Z. I. (2014). Why the Monte Carlo method is so important today. *WIREs Computational Statistics*, 6(6), 386–392. <https://doi.org/10.1002/wics.1314>
12. Kucherenko, S., Tarantola, S., & Annoni, P. (2012). Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications*, 183(4), 937–946. <https://doi.org/10.1016/j.cpc.2011.12.020>
13. Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters*, 3, 18–23. <https://doi.org/10.1016/j.mfglet.2014.12.001>
14. Lee, J., Davari, H., Singh, J., & Pandhare, V. (2018). Industrial artificial intelligence for Industry 4.0-based manufacturing systems. *Manufacturing Letters*, 18, 20–23. <https://doi.org/10.1016/j.mfglet.2018.09.002>
15. Matzka, S. (2020a). Explainable artificial intelligence for predictive maintenance applications. In *2020 Third International Conference on Artificial Intelligence for Industries (AI4I)* (pp. 69–74). IEEE. <https://doi.org/10.1109/AI4I49448.2020.00023>
16. Matzka, S. (2020b). AI4I 2020 Predictive Maintenance Dataset. UCI Machine Learning Repository. <https://doi.org/10.24432/C5HS5C>
17. Montero Jimenez, J. J., Schwartz, S., Vingerhoeds, R., Grabot, B., & Salaün, M. (2020). Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics. *Journal of Manufacturing Systems*, 56, 539–557. <https://doi.org/10.1016/j.jmsy.2020.07.008>
18. Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning* (pp. 625–632). Association for Computing Machinery. <https://doi.org/10.1145/1102351.1102430>
19. Plischke, E., Borgonovo, E., & Smith, C. L. (2013). Global sensitivity measures from given data. *European Journal of Operational Research*, 226(3), 536–550. <https://doi.org/10.1016/j.ejor.2012.11.047>
20. Ran, Y., Zhou, X., Lin, P., Wen, Y., & Deng, R. (2019). A survey of predictive maintenance: Systems, purposes and approaches. arXiv. <https://arxiv.org/abs/1912.07383>
21. Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3), Article e0118432. <https://doi.org/10.1371/journal.pone.0118432>
22. Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance-based sensitivity analysis of model output: Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2), 259–270. <https://doi.org/10.1016/j.cpc.2009.09.018>
23. Saracco, P., & Pia, M. G. (2013). An exact framework for uncertainty quantification in Monte Carlo simulation. arXiv. <https://arxiv.org/abs/1311.5221>
24. Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data-driven smart manufacturing. *Journal of Manufacturing Systems*, 48, 157–169. <https://doi.org/10.1016/j.jmsy.2018.01.006>
25. Wei, P., Lu, Z., & Song, J. (2015). Variable importance analysis: A comprehensive review. *Reliability Engineering & System Safety*, 142, 399–432. <https://doi.org/10.1016/j.ress.2015.05.018>
26. Yeh, I. (2009). Default of Credit Card Clients. UCI Machine Learning Repository. <https://doi.org/10.24432/C55S3H>
27. Yeh, I.-C., & Lien, C.-H. (2009). The comparison of data mining techniques for the predictive accuracy of the probability of default of credit card clients. *Expert Systems with Applications*, 36(2), 2473–2480. <https://doi.org/10.1016/j.eswa.2007.12.020>
28. Zhu, H., Liu, T., & Zhou, E. (2020). Risk quantification in stochastic simulation under input uncertainty. *ACM Transactions on Modeling and Computer Simulation*, 30(1), 1–24. <https://doi.org/10.1145/3329117>
29. Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in Industry 4.0: A systematic literature review. *Computers & Industrial Engineering*, 150, Article 106889. <https://doi.org/10.1016/j.cie.2020.106889>