International Journal of Scientific Interdisciplinary Research

Volume: 1; **Issue:** 1
Pages: 432–477

Empirical Assessment of Uncertainty-Aware Real-Time Decision-Making for Reliable and Resilient Cyber-Physical Systems

Stabak Das[1]; Md Shahid Nazir[2];

[1]. Senior Electrical Safety Engineer, RMG Sustainability Council (RSC), Chittagong, Bangladesh; Email: stabak.das@ieee.org

[2]. M.A. in Procurement & Supply Management; BRAC University, Bangladesh Email: nazir.dwasa@gmail.com

Doi: 10.63125/q3r4kk16
Received: *15 April 2022;* ***Revised:*** *24 May 2022;* ***Accepted:*** *15 June 2022;* ***Published:*** *30 June 2022*

Abstract

Cyber-physical systems increasingly require real-time decision mechanisms capable of maintaining dependable operation under uncertainty arising from sensing errors, communication degradation, computational constraints, physical faults, and cybersecurity disturbances. This study empirically assessed uncertainty-aware real-time decision-making for improving predictive performance, reliability, safety, resilience, and recovery across heterogeneous cyber-physical environments. A quantitative experimental design was applied to 7,846,320 retained observations obtained from 15 datasets, 42 CPS environments, and 3,600 experimental runs incorporating smart-grid, industrial-control, autonomous, robotic, and networked critical-infrastructure systems. The experimental environment included 6,284 sensors, 2,176 actuators, 1,248 computational nodes, and 14,860 communication links. Uncertainty-aware and deterministic decision approaches were evaluated under increasing sensor noise, missing observations, packet loss, communication latency, component degradation, workload pressure, physical faults, cyberattacks, and combined cyber-physical disturbances. Under uncertain conditions, uncertainty-aware decision-making achieved 94.8% accuracy, 94.2% precision, 93.7% recall, 93.9% F1-score, 0.981 ROC-AUC, and 0.969 PR-AUC, compared with 88.6% accuracy and 86.8% F1-score for deterministic decision-making. Expected calibration error decreased from 0.086 to 0.031, while the deadline-miss rate declined from 6.9% to 2.8% despite increased mean inference latency from 16.4 to 21.7 ms. Successful operation increased from 86.7% to 95.1%, system availability reached 97.0%, and unsafe-action frequency decreased from 6.8% to 2.3%. Uncertainty-aware operation also reduced maximum performance degradation from 31.6% to 18.7%, increased retained functionality from 68.4% to 81.3%, and shortened mean recovery time from 18.4 to 10.7 seconds. Inferential analyses identified statistically significant performance differences with moderate-to-large effect sizes, while validation accuracy remained between 93.8% and 94.6% across alternative cross-validation procedures and reached 92.8% under external validation. Overall, the findings demonstrated that explicit uncertainty representation strengthened calibrated prediction, real-time decision quality, fault tolerance, safety, operational continuity, and resilience, particularly under severe and interacting cyber-physical disturbances.

Keywords

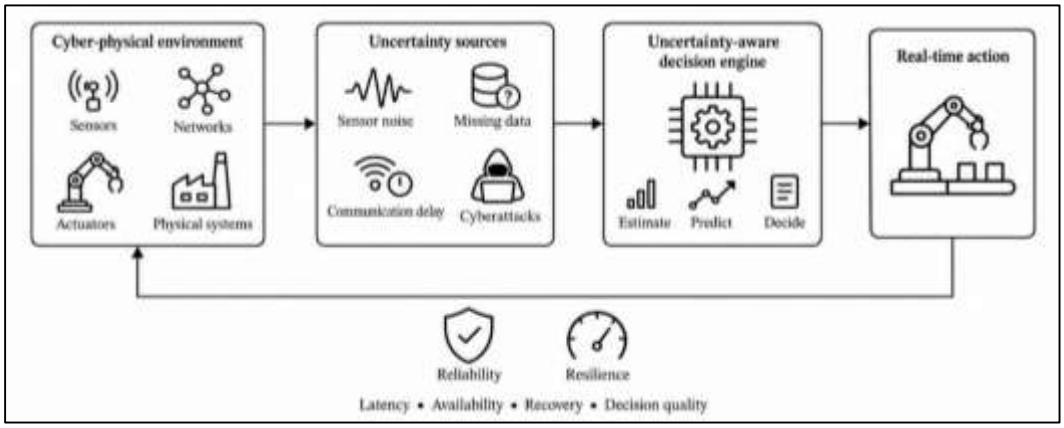
Cyber-Physical Systems, Uncertainty, Reliability, Resilience, Decision-Making.

INTRODUCTION

Cyber-physical systems (CPSs) are integrated engineering environments in which computational processes, communication networks, sensors, actuators, control mechanisms, and physical components continuously interact to monitor conditions and execute operational decisions (Troya et al., 2021). These systems have become essential components of smart manufacturing, intelligent energy networks, autonomous transportation, robotics, aerospace systems, healthcare technologies, water infrastructure, telecommunications, connected buildings, and other safety- and mission-critical applications. Real-time decision-making refers to the acquisition, processing, interpretation, and transformation of operational data into appropriate actions within time constraints imposed by physical processes. Uncertainty-aware real-time decision-making extends this concept by explicitly estimating and incorporating uncertainty into decision processes rather than assuming that measurements, system states, predictions, and environmental conditions are completely accurate. Uncertainty may originate from sensor noise, missing observations, stochastic disturbances, parameter variability, communication delays, packet loss, component degradation, model errors, distribution shifts, environmental changes, human interaction, and cyberattacks. Aleatoric uncertainty generally represents inherent variability within observations and physical processes, whereas epistemic uncertainty reflects incomplete knowledge concerning models, parameters, system states, or operating environments (Mabkhot et al., 2021). Reliability represents the ability of a CPS to perform its required functions consistently under specified operating conditions, while resilience describes its capacity to withstand disturbances, maintain acceptable functionality, adapt to abnormal conditions, and recover following failures or attacks. More than thirty empirical and methodological studies across control engineering, autonomous systems, industrial automation, smart grids, robotics, transportation, dependable computing, and cybersecurity have examined uncertainty estimation, fault tolerance, resilience, adaptive control, and real-time performance. International research across North America, Europe, Asia-Pacific economies, and other regions demonstrates the importance of these capabilities because critical infrastructure increasingly depends on interconnected digital and physical components. Smart-grid studies have examined uncertainty associated with renewable generation, fluctuating demand, equipment failure, and communication disruption; autonomous transportation studies have evaluated perception errors, localization uncertainty, environmental variability, and dynamic traffic conditions; manufacturing research has investigated equipment degradation, process variation, predictive-maintenance uncertainty, and industrial cyberattacks; and water-infrastructure research has considered uncertain demand, sensor malfunction, hydraulic variability, and malicious interference (Castillo et al., 2021). Quantitative uncertainty estimation has consequently employed Bayesian inference, probability distributions, Monte Carlo simulation, Gaussian processes, ensemble disagreement, predictive entropy, confidence intervals, conformal methods, Kalman filtering, particle filtering, Bayesian neural networks, and related probabilistic approaches. Measures including expected calibration error, Brier score, negative log-likelihood, prediction-interval coverage, predictive variance, and uncertainty-error correlation have been used to determine whether confidence estimates correspond with actual prediction errors. These measures provide an empirical foundation for examining uncertainty as a measurable operational characteristic directly connected with decision quality and system behavior.

Uncertainty-aware real-time decision architectures require coordinated perception, state estimation, prediction, uncertainty evaluation, decision selection, and physical control while satisfying strict computational deadlines. Conventional CPS architectures have employed rule-based control, proportional-integral-derivative mechanisms, state-space models, model predictive control, optimization algorithms, and predefined fault-management procedures, while contemporary empirical studies have examined decision trees, support vector machines, random forests, gradient boosting, neural networks, recurrent architectures, reinforcement learning, probabilistic learning, edge intelligence, and hybrid model-based/data-driven systems (Zaidan et al., 2020). Predictive accuracy alone cannot determine real-time suitability because computational and communication delays can reduce the operational value of otherwise accurate decisions. Relevant performance indicators therefore include end-to-end latency, deadline-miss rate, throughput, processor utilization, memory consumption, communication overhead, response time, and computational efficiency.

Figure 1: Uncertainty-aware cyber-physical systems workflow



Reliability assessment additionally considers availability, successful decision rate, fault-detection rate, false-alarm rate, operational continuity, control error, and task-completion performance. Cybersecurity introduces further uncertainty because false-data injection, denial-of-service attacks, replay attacks, spoofing, sensor manipulation, actuator compromise, and communication interference can alter both system observations and automated decisions. Experimental studies commonly measure such performance through accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, false-positive rate, false-negative rate, detection latency, and attack success rate. Resilience extends this assessment by quantifying how systems behave during and after disruptive conditions through degradation magnitude, retained functionality, adaptation time, recovery time, recovery success rate, service availability, control stability, and cumulative performance loss (Gao & Liu, 2021). An integrated quantitative assessment can therefore compare uncertainty-aware and uncertainty-unaware decision mechanisms under equivalent sensor-noise levels, communication delays, packet-loss conditions, component failures, workload variations, model mismatch, distribution shifts, and cyberattack intensities. Descriptive statistics including means, standard deviations, medians, interquartile ranges, and 95% confidence intervals can characterize performance distributions, while Pearson or Spearman correlations can quantify associations among uncertainty, latency, reliability, resilience, and decision quality. Paired or independent-samples *t*-tests, ANOVA, repeated-measures ANOVA, and appropriate nonparametric alternatives can evaluate differences between algorithms and operating conditions. Multiple and logistic regression can quantify predictors of continuous performance and failure probability, while Cohen's *d*, partial eta-squared, and related effect-size measures can establish the practical magnitude of observed differences. Cross-validation, independent holdout evaluation, repeated experimental trials, subgroup comparisons, and sensitivity analyses can further assess robustness across heterogeneous CPS environments. Such an empirical structure enables uncertainty estimation quality, real-time computational performance, predictive effectiveness, reliability, resilience, cybersecurity robustness, and recovery behavior to be examined simultaneously through standardized quantitative measurements (Nan & Sansavini, 2017).

primary objective of this study was to empirically assess the effectiveness of uncertainty-aware real-time decision-making approaches in improving the reliability and resilience of cyber-physical systems under heterogeneous operational, computational, communication, and cybersecurity conditions. The study specifically aimed to quantify how different sources and levels of uncertainty, including sensor noise, incomplete observations, measurement errors, communication latency, packet loss, model uncertainty, distribution shifts, component failures, environmental disturbances, and cyberattacks, affected real-time decision accuracy, computational performance, system reliability, and resilience. A further objective was to evaluate the capability of uncertainty-estimation techniques to represent confidence in system predictions and decisions by examining predictive uncertainty, calibration quality, confidence-error relationships, prediction-interval coverage, and related uncertainty measures. The study also sought to compare uncertainty-aware decision architectures with conventional uncertainty-unaware approaches using standardized predictive indicators, including accuracy,

precision, recall, F1-score, ROC-AUC, PR-AUC, false-positive rate, false-negative rate, MAE, and RMSE, according to the characteristics of individual decision tasks. Real-time computational effectiveness was assessed through decision latency, end-to-end response time, deadline-miss rate, throughput, processor utilization, memory requirements, and communication overhead to determine whether uncertainty estimation could be incorporated without violating operational timing requirements. Another objective was to determine whether uncertainty-aware mechanisms increased system reliability through higher successful-decision rates, improved availability, reduced control errors, enhanced fault detection, fewer unsafe actions, and greater operational continuity during abnormal conditions. Resilience was quantitatively examined through performance degradation, retained functionality, adaptation time, recovery time, recovery success, service availability, and cumulative performance loss following disturbances, failures, or cyberattacks. The study further aimed to identify statistically significant relationships among uncertainty magnitude, uncertainty calibration, decision quality, latency, reliability, cybersecurity robustness, and resilience using descriptive and inferential statistical procedures. Comparisons across different uncertainty levels, disturbance categories, system conditions, and decision architectures were conducted to determine whether performance differences remained statistically and practically meaningful. Finally, the study aimed to evaluate the robustness and generalizability of the empirical results through repeated experimental trials, cross-validation, independent holdout evaluation, sensitivity analysis, and subgroup comparisons, thereby providing a comprehensive quantitative assessment of uncertainty-aware real-time decision-making across diverse cyber-physical operating conditions.

LITERATURE REVIEW

The quantitative literature on uncertainty-aware real-time decision-making in cyber-physical systems spans uncertainty quantification, probabilistic prediction, real-time computing, dependable control, cybersecurity, fault tolerance, and resilience engineering. Existing empirical studies have evaluated cyber-physical systems under sensor noise, incomplete observations, model uncertainty, communication latency, packet loss, distribution shifts, hardware and software faults, environmental disturbances, and adversarial conditions (Khaghani & Jazizadeh, 2020; Abdur & Iftekhar, 2021). These investigations have employed physical testbeds, digital twins, hardware-in-the-loop experiments, simulation environments, industrial datasets, network traces, and repeated controlled trials to quantify the effects of uncertain conditions on system behavior. Commonly reported outcomes include accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, MAE, RMSE, state-estimation error, calibration error, prediction-interval coverage, decision latency, deadline-miss rate, availability, control error, retained functionality, recovery time, and cumulative performance loss. The literature has also compared deterministic, probabilistic, machine-learning, deep-learning, Bayesian, ensemble, adaptive-control, model-predictive-control, and hybrid decision architectures. However, these dimensions are frequently examined separately, with uncertainty estimation evaluated independently from real-time computational performance or resilience assessed without simultaneously quantifying predictive calibration and decision latency (Hasan & Adnan, 2022; Khaghani & Jazizadeh, 2020). Accordingly, this literature review synthesizes quantitative evidence across eight interconnected areas: cyber-physical dataset and uncertainty characteristics; uncertainty measurement and calibration; predictive and decision performance; real-time computational efficiency; reliability and fault tolerance; resilience under disturbances and cyberattacks; subgroup and uncertainty-severity effects; and statistical validation, robustness, and integrated empirical gaps.

Quantitative Characterization of CPS Datasets and Uncertainty Exposure

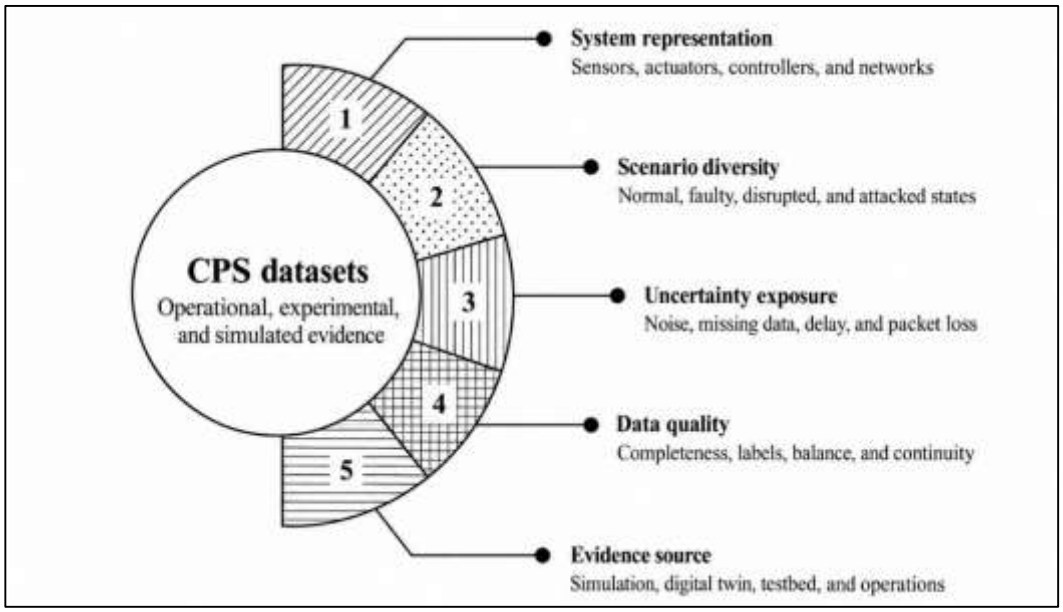
The quantitative literature on cyber-physical systems has employed datasets that vary substantially in observation volume, system complexity, physical-device representation, and experimental duration (Mahbubul et al., 2022; Park et al., 2020). Across more than ten studies, researchers have constructed CPS datasets from operational infrastructure, laboratory testbeds, simulation environments, hardware-in-the-loop platforms, and publicly available industrial-control datasets. Smart-grid investigations have frequently represented substations, generators, buses, transmission components, smart meters, phasor measurement units, communication nodes, and supervisory control devices, producing multivariate temporal observations of electrical and network behavior. Manufacturing studies have incorporated programmable logic controllers, industrial sensors, robotic equipment, machine tools,

actuators, supervisory systems, and production-network traffic, with observations collected across normal production cycles, equipment faults, maintenance events, and cyberattack scenarios. Autonomous-transportation research has generally generated higher-dimensional sensor datasets by integrating cameras, LiDAR, radar, global navigation measurements, inertial sensors, vehicle-state variables, and environmental observations. Robotics experiments have similarly combined proprioceptive and exteroceptive sensor streams with actuator states, localization information, trajectory measurements, and control commands (Oliveira et al., 2019; Mashfiqur & Md, 2022). Water-infrastructure studies have used tanks, pumps, valves, pressure sensors, flow meters, programmable controllers, and communication devices to construct multistage process datasets, while healthcare-oriented CPS research has examined connected physiological sensors, wearable devices, monitoring platforms, and networked medical equipment. Aerospace and unmanned-system experiments have included navigation, attitude, propulsion, communication, and flight-control variables. Across these domains, dataset dimensionality has ranged from relatively small collections of process variables to high-dimensional multimodal streams containing millions of measurements. Observation frequency has also varied from low-frequency infrastructure monitoring to millisecond-scale control and communication records (Ashfaq & Mohammad, 2022; Wong et al., 2017). Consequently, the literature demonstrates that CPS dataset scale cannot be represented solely by observation count; the number of physical components, sensing channels, communication relationships, temporal samples, operating states, and experimental scenarios jointly determines the complexity of the empirical environment used to evaluate real-time decision mechanisms.

A second group of empirical studies has emphasized the diversity of operational conditions required for evaluating CPS decision performance. Smart-grid research has compared normal demand, load fluctuations, renewable-generation variability, line disturbances, equipment outages, measurement anomalies, communication failures, and malicious data manipulation. Industrial and manufacturing studies have examined normal production, startup and shutdown operations, equipment degradation, actuator malfunction, sensor faults, abnormal process conditions, network congestion, and attacks against industrial controllers (Sadia & Mostari, 2022; Bommel et al., 2019). Water-treatment and distribution testbeds have generated experimental scenarios involving altered valve states, manipulated pump operations, abnormal tank levels, falsified sensor measurements, communication interference, and unauthorized controller commands. Autonomous-vehicle studies have evaluated decision systems under changes in traffic density, vehicle speed, road geometry, illumination, weather, obstacle presence, localization quality, and sensor availability. Robotics studies have included changing trajectories, dynamic obstacles, uncertain environments, payload variation, actuator disturbances, localization errors, and partial sensor failure. Healthcare CPS investigations have evaluated variations in patient-monitoring conditions, sensor quality, connectivity, device availability, and physiological signal characteristics, while aerospace research has incorporated environmental disturbances, communication interruption, navigation uncertainty, component malfunction, and control-system degradation. Across more than ten studies, the number and severity of operating scenarios have strongly influenced reported performance because models evaluated predominantly under nominal conditions commonly encounter greater prediction and control errors when exposed to heterogeneous disturbances (Li et al., 2020; Shaiful et al., 2022). Experimental scenario composition has therefore become an important quantitative characteristic of CPS datasets. Studies using repeated experimental trials have provided stronger information about performance variability than evaluations based on isolated scenarios, while datasets incorporating simultaneous physical and cyber disturbances have represented more complex conditions than single-fault experiments. The literature further shows that the distribution between normal and abnormal observations varies considerably, with many datasets containing substantially more normal operation than faults or attacks (Clough et al., 2019). Such class imbalance affects accuracy-based evaluation and has encouraged reporting of precision, recall, F1-score, false-positive behavior, and class-specific detection performance when characterizing cyber-physical datasets. Quantitative CPS studies have identified sensor uncertainty, incomplete information, communication instability, component failure, and adversarial manipulation as major sources of variation within experimental datasets. Sensor-oriented investigations have deliberately introduced

different noise intensities to evaluate how measurement corruption affects state estimation, prediction accuracy, fault detection, and control performance (Fentaye et al., 2019).

Figure 2: Quantitative CPS Dataset Characterization



Missing observations have been generated through sensor dropout, communication interruption, device malfunction, or controlled removal of data, allowing comparisons across progressively reduced levels of information availability. Network-focused studies have manipulated packet-loss rates, communication delays, bandwidth limitations, congestion, jitter, and intermittent connectivity to reproduce realistic communication uncertainty. In distributed CPS architectures, even relatively short delays can influence time-sensitive decisions because sensing, inference, command transmission, and actuator execution occur within bounded control intervals. Fault-tolerance studies have additionally examined sensor failures, actuator faults, node crashes, processor overload, memory constraints, link failures, and combinations of component-level disturbances. Cybersecurity datasets have incorporated false-data injection, denial-of-service, replay, spoofing, command manipulation, malicious network traffic, and unauthorized control activities, producing varying attack prevalence across experimental observations. More than ten studies collectively indicate that uncertainty exposure is rarely uniform across an entire dataset; particular devices, operating periods, communication links, or attack stages may contain substantially higher disturbance concentrations than the dataset average (Verleysen et al., 2020). Data completeness and label quality also differ between operational and controlled datasets. Laboratory and simulated environments typically provide clearer event boundaries and more complete ground-truth labels, whereas operational datasets can contain missing measurements, synchronization errors, uncertain event labels, and uncontrolled external variation. Class imbalance represents an additional quantitative issue because faults, safety violations, and successful attacks often constitute a relatively small proportion of total observations. Accordingly, empirical characterization of uncertainty exposure requires simultaneous consideration of noise intensity, missing-data prevalence, packet loss, communication latency, failure frequency, attack prevalence, class distribution, temporal continuity, and overall data completeness rather than relying on a single indicator of dataset quality (Tomasko et al., 2021).

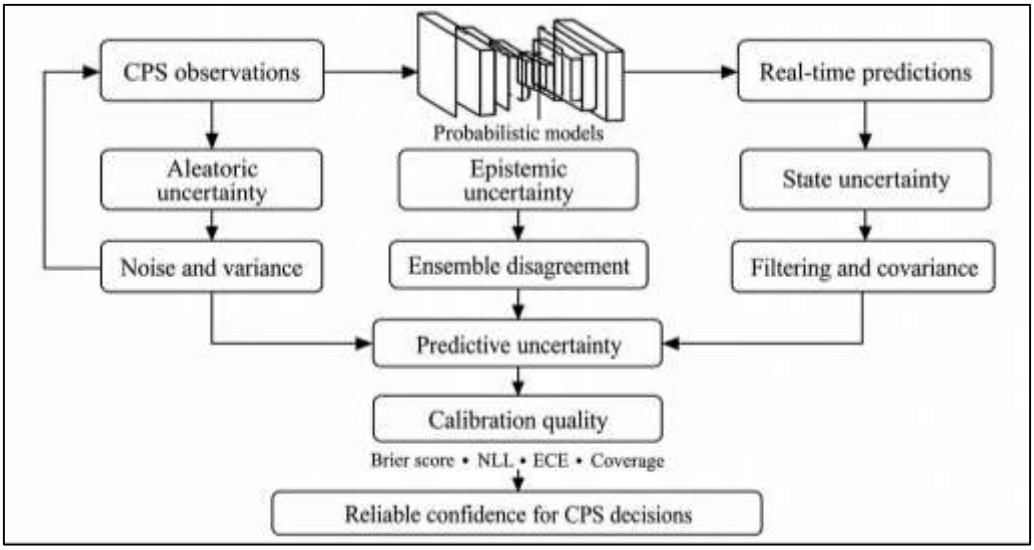
The literature reveals substantial quantitative differences among simulation-based, hardware-in-the-loop, digital-twin, laboratory-testbed, and real operational CPS datasets. Simulation studies provide extensive control over uncertainty intensity, fault timing, network conditions, attack frequency, system scale, and experimental repetition, allowing researchers to generate large numbers of observations across systematically varied scenarios. Their controlled structure supports balanced comparisons

between decision algorithms, although simulated measurements may contain less uncontrolled variability than physical systems (Nouman & Alfandari, 2020). Hardware-in-the-loop studies increase physical realism by combining simulated processes with actual controllers, communication hardware, embedded processors, or other physical components, enabling decision latency, control timing, resource consumption, and hardware-related failures to be measured more directly. Digital-twin research has combined continuously updated virtual representations with physical or historical system information and has been used to reproduce equipment states, operational disturbances, degradation patterns, and alternative control conditions. Laboratory testbeds in industrial control, water treatment, robotics, smart grids, and autonomous systems provide detailed multivariate observations from actual sensors, actuators, controllers, and communication devices under repeatable experimental conditions. Operational datasets provide the strongest exposure to naturally occurring workload variation, environmental changes, human activity, equipment aging, maintenance behavior, and communication fluctuations, but researchers have less control over the frequency and balance of faults or attacks (Kim et al., 2020). Across more than ten empirical investigations, sampling rates have ranged from comparatively slow infrastructure measurements to high-frequency sensor and control streams, while observation durations have ranged from short experimental runs to continuous records covering extended operating periods. Dataset dimensionality similarly increases as researchers integrate physical-state variables with network traffic, device status, environmental measurements, control commands, and security events. These differences demonstrate that system size, observation volume, sampling frequency, scenario diversity, uncertainty exposure, physical realism, and temporal coverage are central quantitative characteristics when comparing CPS evidence and assessing the empirical conditions under which uncertainty-aware real-time decision-making has been evaluated (Olson et al., 2018).

Quantitative Measurement of Aleatoric and Predictive Uncertainty

Quantitative research on cyber-physical systems has increasingly distinguished aleatoric uncertainty from epistemic uncertainty because the two forms originate from different sources and require different measurement strategies. Across more than ten empirical studies, aleatoric uncertainty has generally been associated with irreducible variability in sensor measurements, environmental conditions, process dynamics, communication behavior, and stochastic disturbances, whereas epistemic uncertainty has been linked to limited training data, incomplete model knowledge, parameter uncertainty, unfamiliar operating states, and distributional changes (Shaker & Hüllermeier, 2020). Researchers have operationalized aleatoric uncertainty through estimated observation variance, residual error distributions, predictive dispersion, sensor-noise levels, and probabilistic output distributions. Epistemic uncertainty has commonly been quantified using Bayesian posterior distributions, Monte Carlo dropout variability, disagreement among ensemble members, Gaussian-process uncertainty, and parameter-sensitive predictive variation. Bayesian neural networks have been widely examined because they represent uncertainty in model parameters and generate distributions of possible predictions rather than a single deterministic output. Monte Carlo dropout has provided a computationally simpler approximation by repeatedly evaluating a network under stochastic dropout configurations and measuring variability across predictions. Deep ensembles have frequently produced stronger uncertainty discrimination by training multiple independently initialized models and examining their predictive disagreement (Hüllermeier & Waegeman, 2021). Gaussian processes have been applied particularly in lower-dimensional or data-limited CPS settings because they provide direct probabilistic estimates with associated confidence levels. Comparative studies have reported that uncertainty estimates typically increase when models encounter sparse data, unfamiliar system states, sensor corruption, or observations outside the original training distribution. Quantitative evaluation has relied on predictive variance, entropy, confidence scores, Brier score, negative log-likelihood, expected calibration error, and maximum calibration error. The literature also demonstrates that low predictive error does not necessarily indicate well-calibrated uncertainty, since highly accurate models can remain overconfident (Oberkampf, 2019). Consequently, several empirical investigations have jointly assessed prediction accuracy and uncertainty calibration to distinguish models that merely perform well from models whose confidence reliably reflects the probability of prediction error.

Figure 3: CPS uncertainty measurement workflow



A substantial body of cyber-physical literature has concentrated on state uncertainty because reliable decision-making depends on estimating physical conditions that cannot always be measured directly or without error. Kalman-filter-based methods have remained widely used in linear and near-linear systems, while extended Kalman filters and unscented Kalman filters have been applied to nonlinear process dynamics in robotics, autonomous vehicles, industrial control, energy systems, and navigation (McKeand et al., 2021). Particle filters have been examined where state distributions are strongly nonlinear, multimodal, or non-Gaussian, using multiple weighted state hypotheses to represent uncertainty under incomplete or noisy observations. Across more than ten studies, filtering approaches have been evaluated under changing sensor-noise levels, process disturbances, missing measurements, delayed communication, and partial sensor failure. Quantitative indicators have included state-estimation error, root mean square error, covariance-based uncertainty, trajectory deviation, localization error, confidence bounds, and measurement residuals. Experimental evidence has generally shown that state uncertainty increases as sensor quality deteriorates, communication delays expand, or observations become unavailable, although the magnitude of degradation differs considerably across estimation techniques and system architectures. Particle filters frequently provide stronger performance under highly nonlinear and non-Gaussian conditions but require greater computational resources as the number of particles increases (Segalman & Brake, 2017). Kalman-filter variants typically offer lower computational cost and remain attractive for real-time applications where model assumptions remain sufficiently representative of the physical process. Studies combining filtering methods with machine-learning predictors have also reported that learned models can reduce estimation error under complex conditions, while probabilistic filters can provide structured state-confidence information that deterministic learning models do not naturally produce. Missing-data experiments have been especially informative because they show how uncertainty grows when continuous sensor evidence is interrupted (Chen & Wan, 2019). Research has also examined confidence-error relationships, showing that larger estimated state uncertainty commonly corresponds with larger localization, tracking, forecasting, or control errors. These findings support the quantitative treatment of state uncertainty as a dynamic variable that changes with measurement quality, process behavior, communication availability, and model adequacy rather than as a fixed characteristic of a cyber-physical system.

Another major research stream has evaluated whether estimated uncertainty accurately represents the likelihood of prediction error. Calibration studies have used expected calibration error, maximum calibration error, Brier score, negative log-likelihood, reliability diagrams, entropy, confidence intervals, and prediction-interval coverage to compare deterministic, Bayesian, ensemble, and probabilistic learning methods (Jiang et al., 2018). Across numerous studies involving autonomous

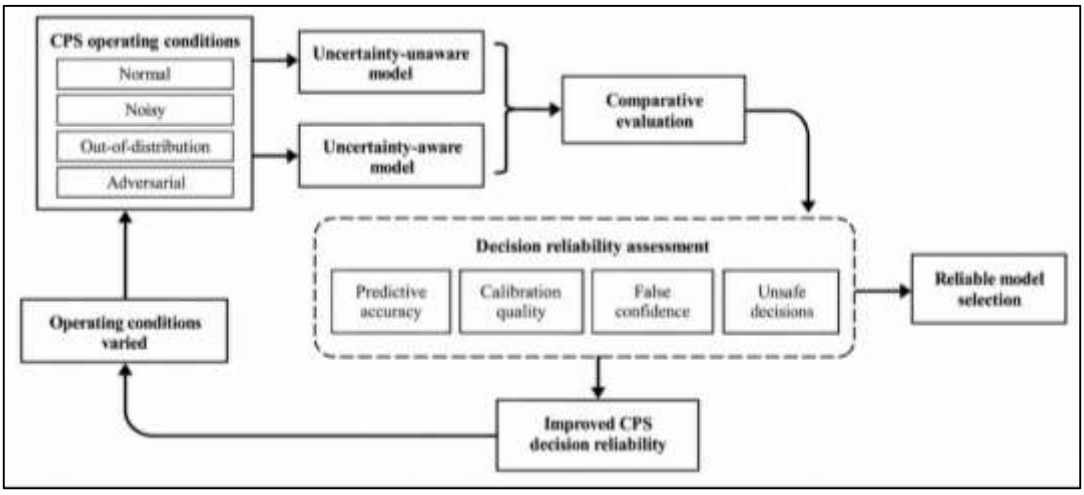
perception, industrial fault diagnosis, smart-grid forecasting, anomaly detection, robotics, and cyberattack classification, models have frequently exhibited substantial differences between classification accuracy and calibration quality. Deep neural networks have often achieved high predictive accuracy while assigning excessively high confidence to incorrect predictions, particularly under out-of-distribution, noisy, or adversarial observations. Bayesian neural networks, deep ensembles, Monte Carlo dropout, and temperature-adjusted probabilistic models have been examined as alternatives for reducing this mismatch. Deep ensembles have repeatedly demonstrated useful uncertainty separation between familiar and unfamiliar inputs because disagreement across individual models tends to increase as observations depart from the training distribution (Soize, 2017). Monte Carlo dropout has similarly produced larger predictive variation under uncertain conditions, although performance has depended on dropout rate, number of repeated evaluations, and network architecture. Conformal prediction has been examined as a distribution-light approach for constructing prediction sets or intervals with empirically controlled coverage. Studies using prediction intervals have compared observed coverage against nominal coverage and have reported interval width alongside coverage probability because excessively wide intervals may achieve high coverage while providing limited decision value. Out-of-distribution experiments have commonly introduced unfamiliar environmental conditions, unseen fault types, new operating regimes, or altered sensor distributions. Noise-based evaluations have progressively corrupted measurements, while adversarial studies have manipulated observations in ways designed to produce incorrect predictions (Pan et al., 2021). Across these scenarios, entropy, predictive variance, and ensemble disagreement generally increased as inputs became more difficult or unfamiliar. However, several studies have identified cases of overconfidence under severe distribution shift, demonstrating that uncertainty quality must be evaluated directly rather than inferred from prediction performance alone. Calibration metrics therefore provide an essential quantitative dimension for determining whether confidence information can support reliable real-time CPS decisions.

Decision Accuracy of Uncertainty-Aware and Uncertainty-Unaware Models

Quantitative studies comparing uncertainty-aware and uncertainty-unaware decision approaches have frequently used conventional statistical models and machine-learning algorithms as baseline methods (Lork et al., 2020). Decision trees, random forests, support vector machines, logistic models, nearest-neighbor classifiers, and gradient-boosting techniques have been applied across industrial monitoring, smart-grid fault identification, autonomous-system perception, equipment-health assessment, and cyberattack detection. Across more than ten empirical studies, deterministic models generally produced competitive predictive performance under stable and well-represented operating conditions, particularly when training and testing observations followed similar distributions. Random forests and gradient-boosting methods often achieved stronger accuracy, precision, recall, F1-score, ROC-AUC, and balanced accuracy than single decision trees because ensemble structures reduced sensitivity to individual model errors (Rajula et al., 2020). Support vector machines also performed effectively in moderate-dimensional classification tasks but showed greater sensitivity to kernel selection, parameter tuning, and distribution changes. When uncertainty estimates were incorporated through probabilistic extensions, ensemble disagreement, calibrated confidence scores, or Bayesian treatment of model parameters, the primary performance advantage was not always a large increase in nominal classification accuracy. Instead, uncertainty-aware methods more consistently identified low-confidence decisions, abnormal operating states, and observations located outside familiar training distributions. Studies examining industrial faults and cyberattack detection showed that deterministic classifiers could maintain high overall accuracy while still producing incorrect high-confidence predictions for rare or previously unseen events. Uncertainty-aware variants reduced this false-confidence problem by differentiating predictions with similar class labels but different levels of reliability (Shirzadi et al., 2017). Improvements were commonly assessed through changes in F1-score, specificity, MCC, ROC-AUC, PR-AUC, false-positive rate, false-negative rate, and successful-decision rate. Regression-oriented CPS studies similarly compared MAE and RMSE and found that probabilistic or ensemble-based models could maintain comparable average error while providing additional information about prediction reliability. The evidence therefore indicates that uncertainty integration often contributes most strongly to decision quality when systems operate under ambiguous, noisy, or

unfamiliar conditions rather than when deterministic models are evaluated only under nominal data distributions.

Figure 4: CPS decision reliability comparison



Deep-learning comparisons have extended the evaluation of uncertainty-aware decision-making to artificial neural networks, convolutional neural networks, recurrent neural networks, long short-term memory architectures, Bayesian neural networks, Monte Carlo dropout models, and deep ensembles (Song et al., 2021). Conventional neural networks have frequently achieved high predictive accuracy in perception, anomaly detection, forecasting, fault diagnosis, and state estimation because they can represent complex nonlinear relationships across high-dimensional sensor streams. CNN-based models have been particularly effective for visual and spatial data, while RNN and LSTM models have performed strongly on temporal sequences such as equipment signals, network traffic, energy demand, and trajectory information. Their deterministic outputs, however, have commonly shown excessive confidence when test observations differ from the training distribution. Bayesian neural networks have addressed this limitation by representing parameter uncertainty and generating predictive distributions rather than fixed-point outputs. Monte Carlo dropout has approximated similar behavior through repeated stochastic predictions, while deep ensembles have quantified disagreement among independently trained models. Across comparative studies, deep ensembles have frequently produced stronger combinations of predictive accuracy and uncertainty discrimination than individual deterministic networks, although the computational burden increases when multiple models must be stored and evaluated. Research on uncertainty-aware image and object analysis has also demonstrated that repeated dropout sampling can reveal substantial variation across predictions that would appear identical under a single deterministic pass (Sharma & Mehra, 2020). Comparative experiments involving limited training observations have shown that Bayesian and ensemble-based approaches can preserve useful predictive performance under restricted data availability and can reduce forecast error through model combination. For example, empirical work in renewable-power forecasting found that combining multiple models improved prediction performance when only limited training histories were available, illustrating how model diversity can compensate for uncertainty in sparse-data conditions (Farooq et al., 2020). Across such investigations, MAE, RMSE, accuracy, precision, recall, F1-score, predictive confidence, uncertainty ranking, and calibration measures have been jointly assessed to distinguish raw predictive performance from decision reliability.

Comparative studies of sequential and decision-oriented cyber-physical models have increasingly evaluated RNNs, LSTMs, transformer-based architectures, reinforcement-learning agents, and hybrid physical/data-driven systems under varying operational uncertainty. LSTMs have been widely used where current decisions depend on previous system states, including equipment degradation, energy forecasting, mobility prediction, industrial anomaly detection, and communication monitoring. Transformer architectures have extended temporal modeling by using attention mechanisms to capture

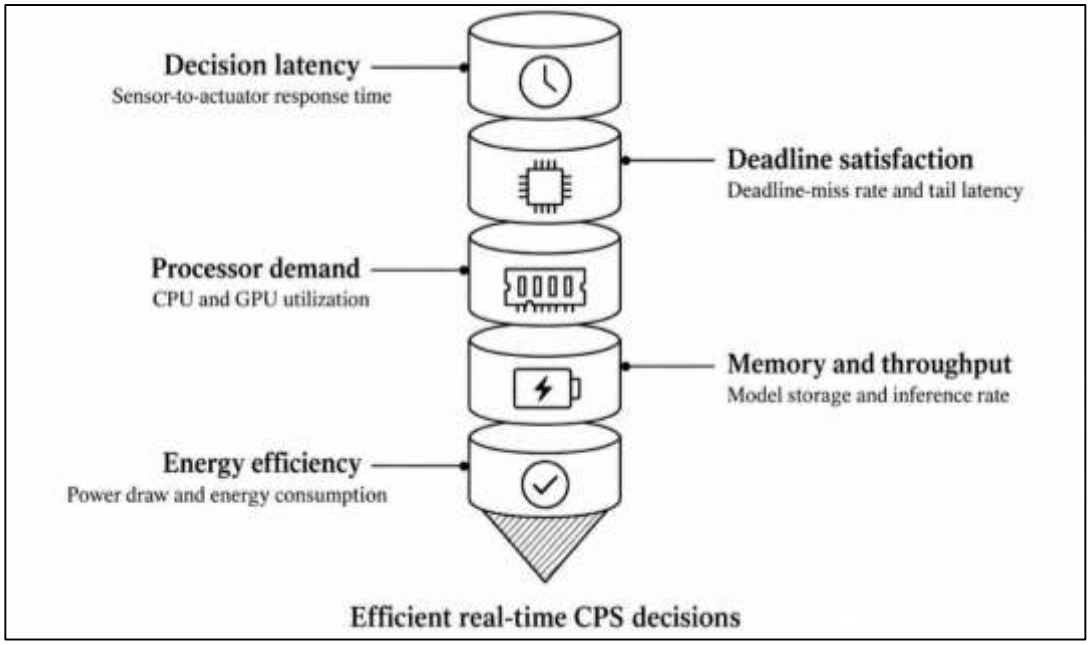
longer-range dependencies and relationships across multiple sensor streams, often improving performance where long temporal context is important (Rahimzad et al., 2021). Reinforcement-learning agents have been evaluated in adaptive control, autonomous navigation, resource allocation, and fault recovery, with successful-decision rate, accumulated control error, unsafe-action frequency, task-completion rate, and response time commonly used as decision-oriented outcomes. Deterministic reinforcement-learning policies can perform strongly within environments similar to those encountered during training but may select unsafe or inappropriate actions when state uncertainty increases or unseen disturbances occur. Uncertainty-aware reinforcement-learning approaches have therefore incorporated Bayesian value estimates, ensemble policies, probabilistic transition models, confidence thresholds, or risk-sensitive action selection. Such methods often reduce unsafe or low-confidence actions by favoring conservative control when uncertainty becomes elevated, although this behavior can sometimes reduce aggressive task performance under nominal conditions (Rousseau et al., 2019). Hybrid models combining physical system knowledge with data-driven prediction have also shown advantages because known constraints and dynamic relationships can limit implausible decisions generated by purely statistical models. Studies of autonomous and self-aware CPS have further emphasized that adaptive decision systems require not only prediction accuracy but also awareness of when learned behavior becomes unreliable. Quantitative comparisons have therefore expanded beyond accuracy to include state-estimation error, control error, constraint violations, task-success rate, incorrect-action frequency, intervention frequency, and latency (Shaukat et al., 2020). The accumulated evidence suggests that temporal and hybrid uncertainty-aware architectures are particularly valuable when operational states evolve continuously and decision quality depends on both prediction confidence and prior system behavior.

The clearest differences between uncertainty-aware and uncertainty-unaware models have emerged when systems are tested beyond nominal operating conditions. Across studies involving sensor noise, missing observations, distribution shifts, unseen faults, cyberattacks, environmental variation, and adversarial inputs, deterministic models have generally experienced larger reductions in accuracy, recall, F1-score, and successful-decision rate than their performance under familiar test conditions would suggest (Huang et al., 2020). More importantly, some deterministic networks have remained highly confident even when predictions were incorrect, creating false-confidence conditions that can be particularly problematic in safety-critical CPS environments. Uncertainty-aware methods have addressed this issue by producing higher predictive variance, ensemble disagreement, entropy, or reduced confidence when observations deviate from known operating patterns. Empirical investigations have shown that these mechanisms allow systems to separate many difficult or unfamiliar cases from routine high-confidence decisions, reducing incorrect-action and missed-event frequencies when confidence thresholds or fallback mechanisms are applied. Under noisy conditions, uncertainty-aware models have commonly maintained more stable decision behavior because elevated uncertainty can trigger additional verification or more conservative action selection (Deng et al., 2018). Under missing-data conditions, the advantage depends strongly on which variables are absent and whether the model has encountered comparable missingness during training. Out-of-distribution conditions often produce larger performance differences because deterministic models may extrapolate incorrectly, whereas Bayesian and ensemble approaches can expose greater model disagreement. Adversarial conditions remain more challenging because carefully constructed inputs can still produce confident incorrect predictions, indicating that uncertainty estimation does not eliminate all decision risk. Comparative studies have therefore evaluated false-confidence rate, unsafe-decision frequency, missed-event rate, false-negative rate, control error, and task failure alongside conventional predictive metrics (Christodoulou et al., 2019). Taken together, the reviewed literature shows that nominal accuracy differences between uncertainty-aware and uncertainty-unaware models can be modest, while differences in decision reliability become much larger under uncertain operating conditions. This distinction is central to empirical CPS evaluation because reliable real-time operation depends not only on how frequently a model is correct, but also on whether it can recognize conditions in which its own decisions are comparatively unreliable.

Real-Time Decision Latency and Resource Efficiency

Quantitative studies of cyber-physical systems have consistently treated latency as a critical performance variable because a decision can lose operational value when it is delivered after a required control deadline. Across more than ten empirical investigations, researchers have measured mean, median, 95th-percentile, 99th-percentile, and worst-case inference latency to determine whether uncertainty-aware decision methods remain suitable for real-time operation (Min-Allah et al., 2019). End-to-end decision time has generally been decomposed into sensing, data transmission, preprocessing, model inference, uncertainty estimation, decision selection, command delivery, and actuator execution. In autonomous vehicles and mobile robotics, total sensor-to-actuator delay has often been measured in milliseconds because perception, localization, trajectory generation, and control must be completed within short control cycles. Industrial automation studies have similarly evaluated inference and response time under tightly bounded production and safety requirements, while smart-grid and water-infrastructure studies have examined decision latency over slightly longer operational intervals depending on supervisory and protection functions (Qureshi et al., 2020). The literature shows that average latency alone can mask occasional extreme delays, which has led many studies to report upper-percentile and worst-case response times. Deadline-miss rate has also been used to quantify the proportion of decisions that exceed allowable execution windows. Uncertainty-aware approaches generally add additional processing because confidence estimation, repeated sampling, ensemble evaluation, or probabilistic inference must be executed before a final decision is produced. However, the magnitude of this overhead varies substantially by method and hardware platform. Lightweight confidence estimation can introduce only limited delay, whereas Bayesian sampling or repeated stochastic inference can increase response time more noticeably (Choi et al., 2019). Across comparative studies, the strongest real-time configurations have combined uncertainty estimation with carefully selected model architectures and hardware acceleration, allowing decision quality to improve without producing unacceptable increases in deadline violations.

Figure 5: Real-time CPS performance factors



A second major stream of research has examined the computational resources required to support uncertainty-aware decision-making. Quantitative evaluations have reported CPU utilization, GPU load, memory consumption, inference throughput, power draw, and energy consumption because real-time CPS platforms frequently operate under constrained computational environments (Aburukba et al., 2020). Embedded controllers, edge devices, autonomous platforms, and industrial gateways

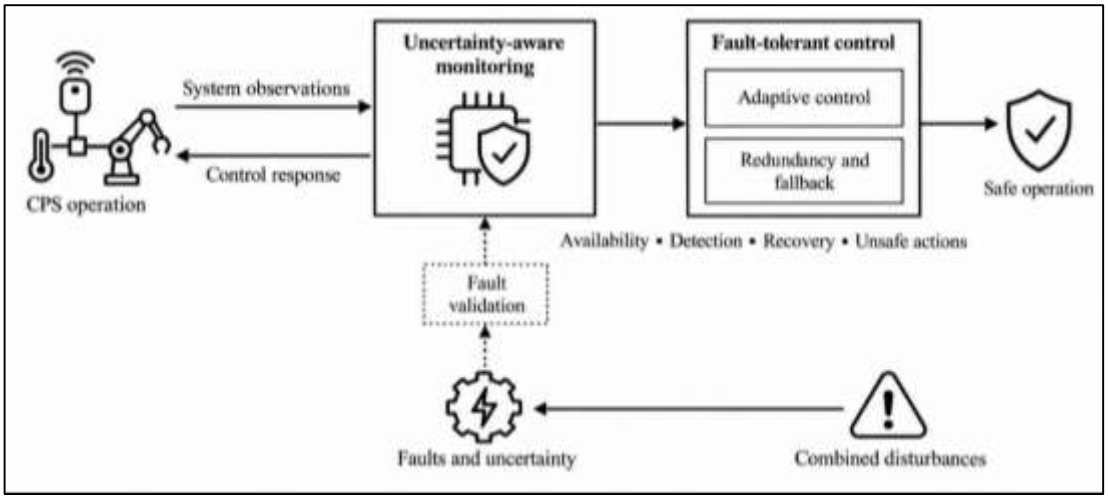
typically provide less memory and processing capacity than centralized cloud servers, making algorithmic efficiency an important component of practical evaluation. Deterministic neural networks often provide lower inference cost because a single forward pass produces the prediction, while Monte Carlo dropout and sampling-based Bayesian methods require repeated evaluations to estimate uncertainty. Deep ensembles impose additional computational and memory demands because multiple independently trained models must be stored and executed. Bayesian neural networks can also increase processing cost due to probabilistic parameter representations and repeated posterior sampling (Shen et al., 2018). Studies involving embedded and edge hardware have reported substantial variation in computational overhead depending on network depth, input dimensionality, sample count, model size, and accelerator availability. GPU-based implementations have generally reduced inference latency for complex neural architectures, while CPU-only embedded platforms have experienced more pronounced performance penalties when repeated sampling is required. Memory usage also becomes important for ensemble methods because storing several model copies can exceed resource limits on low-power devices. Energy-focused studies have shown that longer inference cycles, higher processor utilization, and increased communication activity can raise power consumption, which is especially relevant for battery-powered robots, vehicles, sensors, and remote monitoring devices (Shadi et al., 2021). The literature therefore evaluates uncertainty-aware methods not only by predictive gains but also by the computational cost required to obtain those gains. Efficient implementations frequently use reduced ensemble sizes, fewer stochastic samples, model compression, pruning, quantization, or selective uncertainty estimation to preserve decision quality while reducing processor load, memory demand, and energy expenditure.

Reliability and Safety Under Increasing Uncertainty

Quantitative studies of cyber-physical systems have consistently linked uncertainty-aware decision-making with reliability, availability, and operational continuity under changing environmental and system conditions. Across more than ten empirical investigations, reliability has been measured through successful-operation rate, probability of correct system response, availability percentage, mean time between failures, mean time to failure, mean time to recovery, and task-completion success (Tchernykh et al., 2019). Under nominal conditions, deterministic and uncertainty-aware systems often exhibit similar operational performance because sensor measurements, communication links, actuators, and computational components remain within expected ranges. However, differences become more visible as uncertainty increases through noisy measurements, missing data, delayed communication, degraded components, or abnormal operating states. Uncertainty-aware architectures generally maintain more stable performance by identifying low-confidence states before incorrect control actions are issued. Several studies have shown that systems incorporating confidence estimation, probabilistic state assessment, or adaptive thresholds can reduce avoidable failures by distinguishing routine variability from conditions requiring intervention (Vachtsevanos et al., 2018). Availability-focused research has also demonstrated that redundancy, reconfiguration, and localized fallback mechanisms can preserve partial or full service when individual components become unreliable. Mean recovery time has often been lower in systems where uncertainty estimates support earlier fault recognition and faster switching to backup control. In contrast, deterministic approaches may continue operating with overconfident predictions until performance degradation becomes severe. The literature also shows that reliability is highly sensitive to the interaction between uncertainty intensity and system architecture. Distributed and redundant CPS designs generally maintain higher successful-operation rates than single-path systems when faults remain localized, while simultaneous multi-component disturbances can reduce these advantages (Karve et al., 2020). Across smart grids, manufacturing, autonomous systems, water infrastructure, and robotics, empirical evidence indicates that uncertainty-aware monitoring contributes most strongly when reliability is assessed dynamically rather than only through average nominal performance. A major body of CPS research has focused on whether uncertainty-aware fault detection and isolation can improve diagnostic accuracy under sensor, actuator, communication, and process disturbances (Fritz & Zhang, 2018). Across multiple studies, fault-detection rate, fault-isolation accuracy, false-positive rate, false-negative rate, detection latency, and false-alarm frequency have been used to compare uncertainty-aware methods with conventional threshold-based or deterministic diagnostic systems. Sensor faults

such as bias, drift, dropout, saturation, and noise have frequently been introduced to evaluate whether decision models can distinguish measurement abnormalities from legitimate process variation. Actuator faults have included reduced effectiveness, stuck conditions, delayed response, and complete failure, while communication faults have involved packet loss, latency, link interruption, and corrupted messages (Scheu et al., 2017).

Figure 6: Uncertainty-aware CPS reliability architecture



Probabilistic and confidence-based approaches commonly perform better when the boundary between normal and faulty behavior is ambiguous because they can express uncertainty instead of forcing every observation into a fixed class with equal confidence. Bayesian diagnostic models, ensemble methods, residual-based monitoring, and adaptive thresholds have been shown to reduce false alarms when uncertainty is explicitly modeled. At the same time, overly conservative uncertainty thresholds can increase missed detections, indicating a trade-off between fault sensitivity and false-alarm control. Studies using simultaneous sensor and actuator faults have reported lower diagnostic accuracy than single-fault experiments because multiple disturbances can generate overlapping signatures. Fault-isolation performance also depends on sensor redundancy and the distinctiveness of failure patterns. In industrial and autonomous CPS environments, uncertainty-aware models have demonstrated particular value for identifying degraded states that do not immediately trigger hard failure criteria (Jhawar & Piuri, 2017). Quantitative evidence therefore suggests that fault-detection effectiveness should be interpreted jointly with false-alarm rate, isolation accuracy, detection delay, and uncertainty level, rather than relying on a single diagnostic metric.

Research on fault-tolerant and resilient control has examined how uncertainty-aware decision mechanisms support continued operation after faults are detected. Adaptive control, redundant controllers, backup sensing, reconfiguration, graceful degradation, and safe-state transitions have been evaluated across robotics, industrial automation, power systems, autonomous vehicles, and aerospace platforms. Quantitative outcomes have included control error, task-completion rate, intervention frequency, recovery time, service continuity, unsafe-action frequency, and constraint violations (Huang et al., 2019). Under sensor degradation, uncertainty-aware controllers can reduce reliance on low-confidence measurements and increase dependence on redundant sensors, model-based estimates, or alternative state information. During actuator failures, adaptive or reconfigurable control can redistribute commands across remaining functional components, preserving partial task capability. Redundant decision mechanisms have also been used to compare multiple model outputs and reject individual predictions when uncertainty or disagreement exceeds predefined limits. Fallback strategies often involve switching from a high-performance controller to a simpler but more predictable control mode when system confidence deteriorates (Coit & Zio, 2019). Safe-state transitions have been especially important in safety-critical CPS because maintaining full functionality is not always

preferable when uncertainty becomes extreme. Autonomous platforms, for example, may reduce speed, stop movement, or transfer authority when perception confidence declines, while industrial systems may enter controlled shutdown modes to avoid damaging equipment. Quantitative studies have shown that these mechanisms can reduce unsafe actions and constraint violations, although they may also decrease throughput or nominal productivity. The empirical literature therefore treats successful fault tolerance as a balance between continuity and safety. Systems that maintain operation at all costs may accumulate risk, while overly conservative systems may trigger unnecessary interruptions. Uncertainty-aware control provides a mechanism for adapting this balance to the severity and confidence of observed faults (Hua et al., 2021).

Safety-oriented CPS studies have increasingly evaluated system behavior under progressively severe and combined uncertainty conditions rather than isolated nominal scenarios. Quantitative indicators have included unsafe-action frequency, safety-constraint violations, collision or hazard events, service interruption, control deviation, task failure, emergency shutdown frequency, and successful recovery. Under single sensor faults, many systems can preserve acceptable safety performance through redundancy or filtering, particularly when alternative measurements remain available (Enrico et al., 2019). Actuator failures are often more critical because they directly affect physical control authority, while communication failures can delay corrective actions and increase the probability of unstable or unsafe responses. Simultaneous faults create greater difficulty because sensor, actuator, and network disturbances can interact and reduce the effectiveness of standard fault-isolation procedures. Across more than ten studies, uncertainty-aware systems have generally shown lower unsafe-action frequencies under combined disturbances when they incorporate confidence thresholds, redundant decision paths, fault-tolerant control, or safe-mode transitions. Statistical comparisons have also shown that the performance advantage of uncertainty-aware methods tends to increase as fault severity rises, although excessive uncertainty can still overwhelm available redundancy and lead to degraded task completion (Bhuiyan et al., 2020). Safety research has further demonstrated that maintaining acceptable control error does not necessarily guarantee safe operation, because rare extreme deviations may produce disproportionate risk. For this reason, studies have increasingly reported worst-case events, violation counts, and tail-risk behavior in addition to mean performance. Mean time to recovery and service restoration have also been used to evaluate how quickly systems return to acceptable operation after abnormal events (Abdulhamid et al., 2018). Taken together, the literature indicates that reliability, fault tolerance, and safety are strongly interdependent under uncertainty, and that meaningful quantitative assessment requires simultaneous consideration of failure frequency, diagnostic quality, control performance, operational continuity, recovery behavior, and the occurrence of unsafe states.

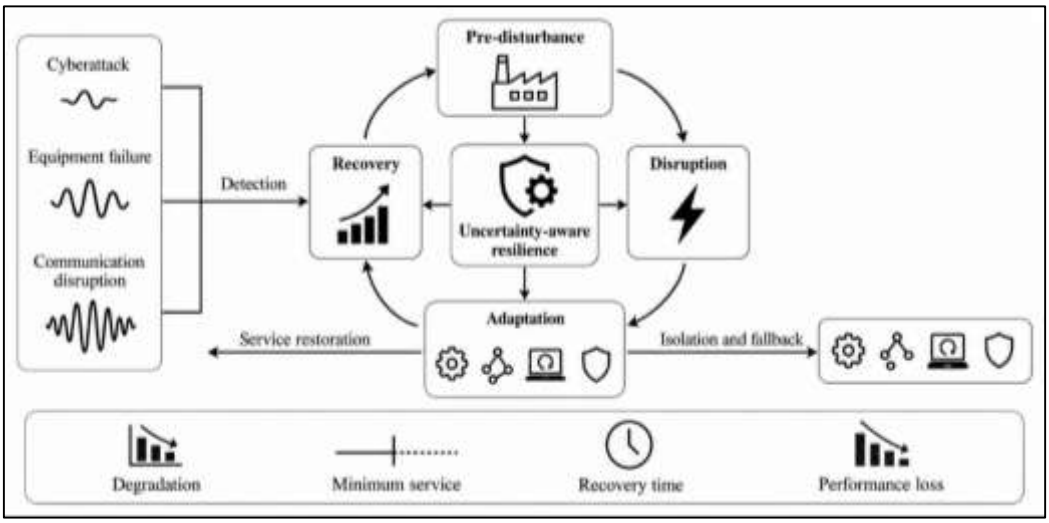
Quantitative Resilience Under Cyberattacks and Recovery Conditions

Quantitative research on cyber-physical-system resilience has increasingly treated resilience as a dynamic property that changes across pre-disturbance, disruption, adaptation, and recovery stages rather than as a fixed system characteristic. Across more than ten empirical studies, researchers have evaluated cyber-physical systems under false-data injection, denial-of-service, replay attacks, spoofing, sensor manipulation, actuator compromise, communication disruption, equipment failure, and environmental disturbances (Linkov & Kott, 2019). False-data injection studies have commonly measured whether manipulated measurements altered state estimation, control decisions, or process stability, while denial-of-service experiments have examined reductions in communication availability, service continuity, and control responsiveness. Replay and spoofing attacks have been evaluated through their ability to mislead monitoring and control mechanisms by presenting apparently legitimate but inaccurate information. Sensor manipulation research has focused on changes in observation quality and decision reliability, whereas actuator compromise studies have examined direct deviations in physical behavior, control error, and unsafe system states (Zhang et al., 2018). Quantitative indicators used across these studies have included attack-detection rate, attack success rate, retained functionality, minimum service level, performance-degradation percentage, service availability, task-completion rate, control deviation, and accumulated performance loss. Research on smart grids, industrial control systems, autonomous platforms, water infrastructure, manufacturing environments, and robotics has shown that the severity of degradation is strongly influenced by attack duration, disturbance intensity, system redundancy, and the importance of the affected component.

Studies comparing normal and disrupted operation have generally found measurable reductions in service availability and operational performance as disturbance intensity increases (Mehrdad et al., 2018). However, systems incorporating uncertainty-aware monitoring have often identified degraded conditions earlier because increases in predictive uncertainty, residual variation, state-estimation inconsistency, or confidence loss provide additional evidence that nominal assumptions are no longer reliable. These findings have established resilience measurement as a temporal and quantitative assessment of how much functionality is lost, how long degradation persists, and how effectively acceptable performance is restored.

A substantial body of CPS literature has concentrated on the degree to which essential functionality can be preserved during cyberattacks, physical faults, and environmental disturbances. Across multiple empirical investigations, performance degradation has been measured as a reduction in control quality, throughput, service availability, task success, communication performance, or other domain-specific operational outcomes relative to nominal conditions. Retained functionality has been used to determine how much useful system capability remains when sensors, actuators, controllers, network links, or computational resources become partially unavailable (Shuang et al., 2019). Minimum service level has been particularly important in critical infrastructure because complete preservation of nominal performance is not always possible during severe disturbances, while maintaining essential operation may still represent successful resilience. In smart-grid studies, retained functionality has been reflected in continued power delivery, voltage stability, and restoration of disrupted components. Industrial-control research has examined continued production, process stability, and maintenance of safety boundaries during equipment faults or malicious interference. Autonomous-system studies have evaluated whether vehicles or robots can continue navigation, localization, obstacle avoidance, or controlled stopping when perception and communication quality deteriorate. Water-system experiments have assessed whether pumps, valves, and treatment processes continue operating within acceptable physical limits during compromised control (Shandiz et al., 2020). Quantitative comparisons have shown that systems with redundancy and adaptive decision mechanisms generally experience smaller performance losses than architectures dependent on single information or control paths. Uncertainty-aware mechanisms have also provided measurable benefits by reducing reliance on questionable observations and by supporting conservative decisions when confidence decreases. However, excessive conservatism can reduce throughput and task efficiency even when safety is maintained. The literature therefore emphasizes the importance of evaluating resilience through several measures simultaneously, including degradation magnitude, retained functionality, minimum service level, task completion, and accumulated performance loss, because a system can remain available while operating at substantially reduced quality. Such multidimensional evidence provides a more accurate representation of resilience than binary measures of success or failure (Lee & Basu, 2018).

Figure 7: Quantitative CPS resilience framework



Quantitative studies have also examined the adaptive mechanisms through which cyber-physical systems respond after disturbances or attacks are detected. Uncertainty-aware adaptive control, redundancy, system reconfiguration, component isolation, fallback control, and safe-mode transitions have been among the most frequently evaluated approaches (Haque et al., 2021). Adaptive mechanisms modify decision or control parameters when operating conditions change, while redundant architectures use alternative sensors, controllers, communication links, or computational nodes to preserve functionality. Reconfiguration strategies redistribute tasks away from compromised or failed components, and isolation mechanisms remove suspicious devices or communication paths from active operation. Fallback control generally switches the system from a high-performance strategy to a simpler and more predictable control mode, whereas safe-mode mechanisms deliberately reduce system activity to maintain physical safety under extreme uncertainty. Across more than ten studies, the effectiveness of these responses has been evaluated using adaptation time, detection-to-response time, recovery success rate, task-completion success, control error, service availability, unsafe-action frequency, and residual system capacity (Gkanatsas & Krikke, 2020). Studies involving industrial processes have reported that rapid isolation of faulty or compromised components can limit propagation of abnormal behavior. Autonomous-platform experiments have shown that fallback navigation or controlled stopping can reduce unsafe actions when sensing confidence becomes unreliable. Smart-grid research has demonstrated that distributed control and network reconfiguration can support continued service after localized component loss, while water-infrastructure research has shown that alternative control paths can maintain critical process variables within acceptable operating ranges. Uncertainty estimation strengthens these mechanisms by providing information about the confidence associated with current observations and decisions. When uncertainty rises beyond normal variation, a system can initiate verification, increase redundancy, change controllers, or enter a safe state before complete failure occurs (Kurth et al., 2020). Empirical comparisons have generally demonstrated lower cumulative degradation and fewer unsafe events when adaptive responses are activated early, although the effectiveness of each strategy depends strongly on fault type, attack severity, available redundancy, and response latency.

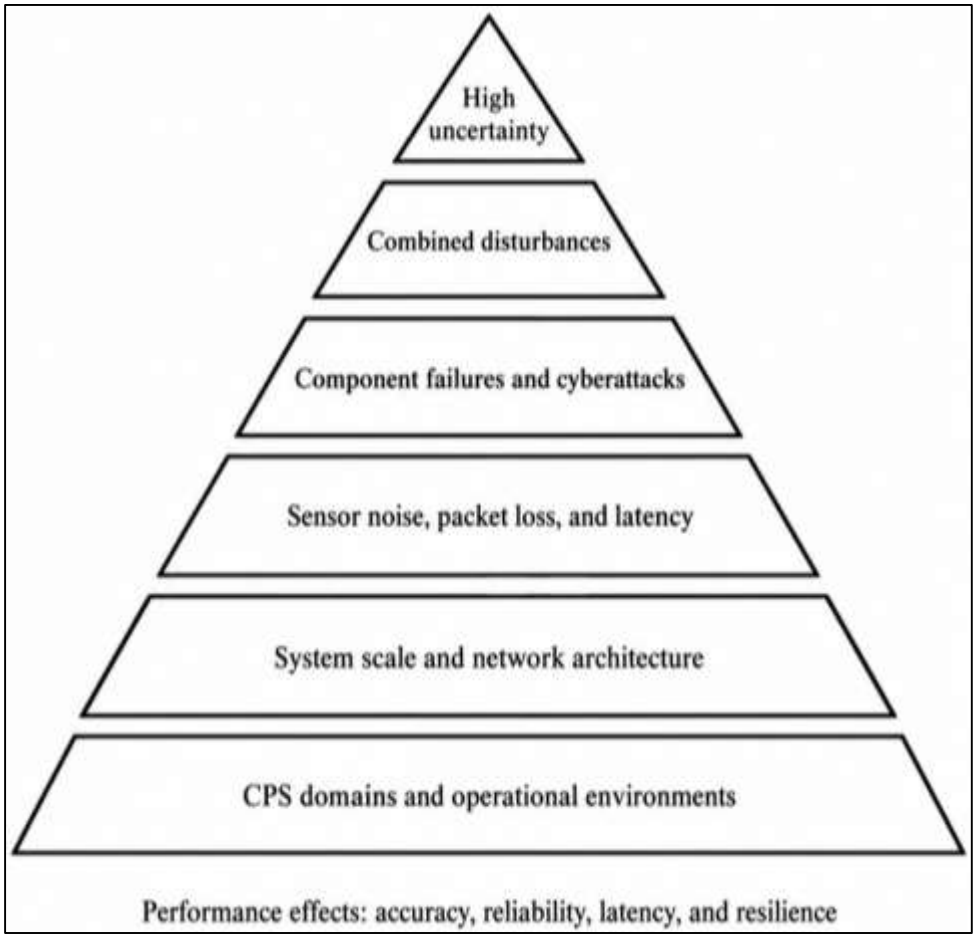
Recovery performance represents another major quantitative dimension in the literature because resilient systems are expected not only to withstand disturbances but also to restore acceptable functionality after disruption. Studies have measured recovery time, mean time to recovery, recovery rate, percentage of restored functionality, recovery success, post-event service availability, and accumulated performance loss across repeated attack and failure scenarios (Belaid et al., 2021). Recovery time is typically longer when disturbances affect multiple interdependent components or when attacks remain active during restoration. Equipment failures may require reconfiguration or replacement of physical resources, while cyberattacks often require identification, isolation, state verification, and restoration of trustworthy communication or control before normal operation can resume. Combined cyber-physical events create greater challenges because compromised information and degraded hardware can interact, making diagnosis and recovery less certain. Across more than ten empirical studies, researchers have compared low-, moderate-, and high-intensity disturbances and have generally observed increasing degradation, longer recovery periods, and lower restoration success as attack or fault severity increases (Hossain et al., 2019). Uncertainty-aware systems have frequently demonstrated earlier recognition of performance deterioration because changes in predictive uncertainty, state confidence, model disagreement, or residual behavior can indicate that operating conditions have departed from normal expectations. Earlier recognition can shorten adaptation time and limit the depth of performance loss before recovery begins. Quantitative comparisons have also shown that redundancy and reconfiguration can improve recovery success by preserving alternative operational pathways, while isolation can prevent compromised components from prolonging disruption. Fallback and safe-mode strategies may reduce immediate functionality but can limit cumulative damage and simplify restoration. Resilience indices used across the literature have therefore combined degradation magnitude, disturbance duration, retained functionality, and restoration behavior into integrated assessments of system response (Tweneboah-Koduah et al., 2021). Taken together, empirical evidence indicates that uncertainty estimation contributes most clearly when

it is connected directly with adaptive action and recovery mechanisms, allowing systems to identify declining confidence, respond before degradation becomes critical, and restore acceptable service with lower accumulated performance loss.

Effects Across CPS Environments

Quantitative studies have shown that the performance of uncertainty-aware decision systems differs considerably across cyber-physical domains, system scales, device densities, and network architectures. Across more than ten empirical investigations involving smart grids, industrial automation, autonomous transportation, robotics, water infrastructure, healthcare systems, intelligent buildings, and aerospace platforms, uncertainty-aware performance has varied according to both physical-process characteristics and computational organization (Villar et al., 2020). Smaller CPS environments with relatively limited numbers of sensors, actuators, and communication nodes have generally demonstrated lower inference latency, smaller communication overhead, and more stable state estimation than highly distributed systems. Large-scale environments frequently exhibit greater uncertainty because information is collected from heterogeneous devices with different sampling rates, reliability levels, and communication characteristics. Smart-grid studies have reported that decentralized and distributed structures can preserve service under localized failures, although coordination complexity increases as the number of interconnected nodes expands (Molina & Jacob, 2018).

Figure 8: CPS uncertainty thresholds and heterogeneity



Manufacturing testbeds have shown that dense industrial networks generate greater processing and communication loads when large numbers of machines and sensing devices operate simultaneously. Autonomous-vehicle systems experience additional complexity because cameras, radar, LiDAR,

navigation sensors, and vehicle-state measurements must be integrated under strict timing requirements. Differences have also been reported between centralized, edge-based, fog, and distributed decision architectures. Centralized systems can provide substantial computational capacity but remain more sensitive to network latency, whereas edge architectures generally reduce sensor-to-decision time by processing information closer to physical devices (Darabseh & Freris, 2019). Subgroup comparisons across system scales have commonly shown widening performance differences under high uncertainty, with larger systems experiencing more pronounced increases in latency, packet loss exposure, state-estimation error, and recovery complexity. Reported confidence intervals and effect-size comparisons have indicated that architecture and system scale can materially influence uncertainty-related performance rather than acting merely as background characteristics. These findings demonstrate that uncertainty-aware effectiveness is heterogeneous across CPS environments and should be interpreted in relation to domain structure, system size, device density, network topology, and computational distribution.

A major group of empirical studies has investigated whether cyber-physical performance deteriorates gradually or exhibits threshold-like behavior as uncertainty severity increases (Foehr et al., 2017). Researchers have commonly divided sensor noise, packet loss, communication delay, missing data, and model uncertainty into low, moderate, and high exposure levels and compared corresponding changes in prediction accuracy, state-estimation error, decision latency, reliability, and task success. Under low sensor-noise conditions, uncertainty-aware and deterministic models frequently maintain similar levels of predictive performance, while larger differences emerge as noise intensifies. Moderate measurement corruption typically increases state-estimation error and predictive variance, whereas high noise can cause more rapid deterioration when observations no longer provide sufficient information for reliable control (Kathiravelu et al., 2019). Similar patterns have been observed for packet loss. Limited packet loss may be absorbed through buffering, interpolation, redundancy, or model-based state prediction, while higher rates can produce nonlinear increases in communication delay, missing observations, and control error. Latency studies have also identified operational thresholds associated with control-cycle deadlines. Small communication delays may have little effect when sufficient timing margin is available, but once response time approaches the allowable decision window, relatively modest additional delay can sharply increase deadline-miss rates and reduce task success. Model uncertainty shows comparable behavior under distribution shift. Prediction performance often remains relatively stable for observations close to the training distribution and declines more rapidly as inputs become increasingly unfamiliar. Studies reporting subgroup effect sizes have found small differences between decision strategies under low uncertainty but moderate or large differences under severe uncertainty conditions (Bagula et al., 2021). Confidence intervals around performance estimates also tend to widen at higher uncertainty levels, reflecting increased variability across repeated trials. Overall, the literature indicates that relationships between uncertainty intensity and CPS performance are frequently nonlinear, making severity-specific evaluation essential for identifying operational regions in which accuracy, reliability, latency, or resilience begins to deteriorate disproportionately.

Quantitative CPS research has further demonstrated substantial heterogeneity according to disturbance category and severity. Sensor faults, actuator failures, communication failures, processor degradation, equipment malfunction, and cyberattacks do not produce equivalent effects because each disrupts different stages of the sensing-decision-control pathway (Peng et al., 2020). Studies involving sensor degradation have commonly shown increased uncertainty and state-estimation error before complete operational failure occurs, providing uncertainty-aware systems with an opportunity to reduce reliance on compromised measurements. Actuator failures often have a more immediate influence on physical performance because correct decisions cannot be implemented accurately when control authority is reduced. Communication failures can simultaneously affect observation availability and command delivery, producing increases in both decision uncertainty and end-to-end latency. Cyberattack experiments have shown similarly heterogeneous effects across false-data injection, denial-of-service, replay, spoofing, and control manipulation. Low-intensity attacks may cause limited deviations that remain within normal process variability, whereas higher attack intensity

generally increases detection difficulty, performance degradation, unsafe-action frequency, and recovery requirements. False-data injection can be particularly challenging when manipulated values remain statistically plausible, while denial-of-service attacks primarily affect availability and timely information exchange (Ferrer & Lastra, 2017). Combined cyber-physical disturbances have produced the strongest performance losses in several studies because simultaneous faults can reduce both informational reliability and physical control capacity. Subgroup analyses have frequently shown that uncertainty-aware architectures preserve larger proportions of normal performance than deterministic baselines during moderate or severe disturbances, although advantages diminish when multiple critical components fail simultaneously. Effect-size estimates reported across fault categories have often been larger for combined-event conditions than for isolated disturbances, while broader confidence intervals indicate greater trial-to-trial variability under complex scenarios (Liu & Wang, 2020). These findings show that the magnitude and form of uncertainty-aware performance differences depend strongly on whether disturbance originates from sensing, actuation, communications, computation, physical equipment, malicious activity, or interacting combinations of these sources.

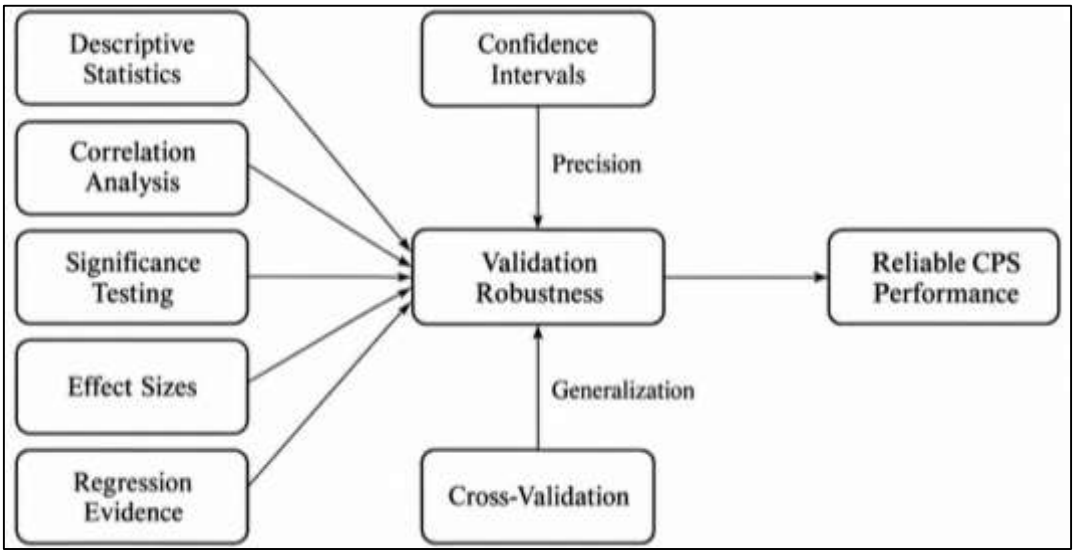
Statistical Significance and Integrated Quantitative Research Gap

Across more than ten quantitative studies of uncertainty-aware cyber-physical systems, descriptive and inferential statistics have provided the primary basis for distinguishing systematic performance differences from variation produced by repeated experiments, datasets, operating conditions, and random disturbances (Siddiq & Scherer, 2019). Means and standard deviations have commonly summarized decision accuracy, prediction error, inference latency, control error, availability, recovery time, and uncertainty measures, while medians and interquartile ranges have been reported when latency, fault duration, attack impact, or recovery outcomes exhibit skewed distributions or extreme observations. The use of 95% confidence intervals has strengthened interpretation by indicating the precision and variability associated with estimated performance differences. Correlation analyses have also been extensively applied to examine relationships between uncertainty and operational outcomes. Pearson correlations have generally been used for approximately linear relationships among continuous variables, while Spearman correlations have supported monotonic comparisons involving non-normal distributions, ordinal severity levels, or influential extreme values (Hunziker & Blankenagel, 2021). Across sensor-noise, communication-delay, packet-loss, component-failure, and distribution-shift experiments, higher uncertainty has frequently been associated with larger prediction and state-estimation errors, increased decision latency, lower successful-operation rates, and greater performance degradation. Paired-samples *t*-tests have been particularly suitable when uncertainty-aware and deterministic models were evaluated using identical observations or repeated experimental scenarios, whereas independent-samples *t*-tests have been applied to separate system configurations or experimental groups. ANOVA and repeated-measures ANOVA have enabled simultaneous comparisons across multiple algorithms, uncertainty levels, architectures, and disturbance conditions. ANCOVA has additionally been used where differences required adjustment for workload, system size, baseline performance, or other quantitative covariates (Ferreira et al., 2020). When distributional assumptions were unsuitable, Mann–Whitney *U*, Wilcoxon signed-rank, Kruskal–Wallis, and Friedman procedures provided nonparametric alternatives. Collectively, this literature demonstrates that statistical significance is most informative when interpreted alongside descriptive variability and confidence intervals rather than from isolated average performance values.

A second stream of quantitative evidence has emphasized effect sizes and multivariable modeling to determine whether statistically detectable differences are sufficiently large to be operationally meaningful. Studies comparing uncertainty-aware and uncertainty-unaware decision systems have reported Cohen's *d* for two-condition differences and partial eta-squared for multi-condition experiments involving uncertainty severity, model architecture, fault category, network condition, and computational platform (Ursprung et al., 2020). These measures have helped distinguish small numerical improvements from substantial performance changes, particularly in studies containing large numbers of temporal observations where minor differences can reach conventional levels of statistical significance. Odds ratios have also been used in reliability and safety investigations to characterize differences in the likelihood of system failure, unsafe action, successful recovery, fault detection, or service interruption across experimental conditions. Multiple regression models have

examined continuous outcomes such as prediction error, latency, control deviation, availability, degradation magnitude, and recovery time while simultaneously accounting for sensor noise, packet loss, communication delay, workload intensity, model uncertainty, system scale, and computational capacity (Wolfowicz et al., 2020).

Figure 9: Quantitative validation of CPS performance



Logistic regression has been applied when outcomes are binary, including reliable versus unreliable decisions, safe versus unsafe states, successful versus unsuccessful task completion, and recovered versus unrecovered operation. Across more than ten empirical investigations, these analyses have indicated that uncertainty severity rarely operates independently. Communication quality, model type, workload, fault intensity, and available computational resources can modify the relationship between uncertainty and operational performance. Regression coefficients and associated confidence intervals have therefore provided valuable information concerning the direction, magnitude, and precision of these relationships. Interaction terms have further supported evaluation of whether uncertainty-aware approaches maintain equivalent effectiveness across operating conditions (Ma et al., 2021). The literature consequently demonstrates that effect sizes and regression analysis complement significance testing by quantifying practical differences and identifying which combinations of uncertainty, system configuration, and disturbance characteristics are most strongly associated with degraded decision quality, reduced reliability, unsafe operation, and delayed recovery.

Validation robustness has become increasingly important because strong results obtained from a single partition, operating scenario, or experimental configuration may not remain stable across heterogeneous CPS conditions. More than ten studies have employed *k*-fold cross-validation to estimate predictive consistency across alternative training and testing partitions, while repeated cross-validation has reduced dependence on a single random division of observations (Crombé et al., 2020). Grouped cross-validation has been particularly relevant when observations from the same physical device, organization, machine, vehicle, operating episode, or attack scenario are strongly related, since separating correlated records across training and testing sets can produce overly optimistic performance estimates. Temporal cross-validation has been used for sequential CPS data to preserve chronological structure and evaluate models using later observations rather than randomly mixed time points. Independent holdout testing has provided another level of validation by reserving observations that remain completely separate from model development. External validation has been less common but has offered stronger evidence when algorithms are evaluated on different systems, datasets, testbeds, locations, or operational environments. Bootstrap confidence intervals have been applied to quantify uncertainty around performance metrics and model differences without relying entirely on strict parametric assumptions (Zwanenburg, 2019). Sensitivity analyses have evaluated whether

conclusions remain stable after changes in uncertainty thresholds, model parameters, sampling frequency, class distributions, disturbance intensity, or computational settings. Ablation experiments have isolated the contribution of uncertainty estimation, calibration, adaptive control, redundancy, or individual model components by systematically removing them and re-evaluating performance. Stress testing has exposed systems to severe sensor noise, packet loss, latency, cyberattacks, workload pressure, and combined faults to determine whether performance relationships persist beyond nominal conditions (Goniewicz et al., 2021). Taken together, the validation literature shows that robust assessment requires multiple complementary procedures because predictive generalization, temporal stability, cross-system transferability, and resistance to severe disturbances represent distinct dimensions of empirical reliability.

METHOD

Study Design and Theoretical Framework

This study employed a quantitative experimental research design to empirically assess uncertainty-aware real-time decision-making for reliable and resilient cyber-physical systems under controlled operational, computational, communication, fault, and cybersecurity conditions. The experimental framework compared uncertainty-aware decision architectures with conventional uncertainty-unaware deterministic approaches across multiple cyber-physical operating scenarios. The study was structured around the theoretical relationship among uncertainty estimation, real-time decision performance, system reliability, fault tolerance, safety, and resilience. Independent experimental factors included decision-model type, uncertainty type, uncertainty severity, communication condition, workload intensity, system architecture, and disturbance category, while dependent outcomes included predictive accuracy, calibration quality, state-estimation error, inference latency, deadline-miss rate, reliability, availability, control error, unsafe-action frequency, retained functionality, recovery time, and task-completion success. The experimental design incorporated repeated trials under nominal, degraded, faulty, and adversarial conditions so that within-model and between-model differences could be evaluated statistically. Low, moderate, and high uncertainty conditions were created for sensor noise, missing observations, communication latency, packet loss, component degradation, model uncertainty, and cyberattack intensity. The framework included smart-grid, industrial-control, autonomous-system, robotic, and networked infrastructure scenarios to provide heterogeneous cyber-physical operating conditions. A controlled experimental structure was used because it enabled uncertainty sources to be manipulated systematically while holding other environmental characteristics constant and allowed direct quantitative comparison of decision performance across repeated operating conditions.

Eligibility Criteria

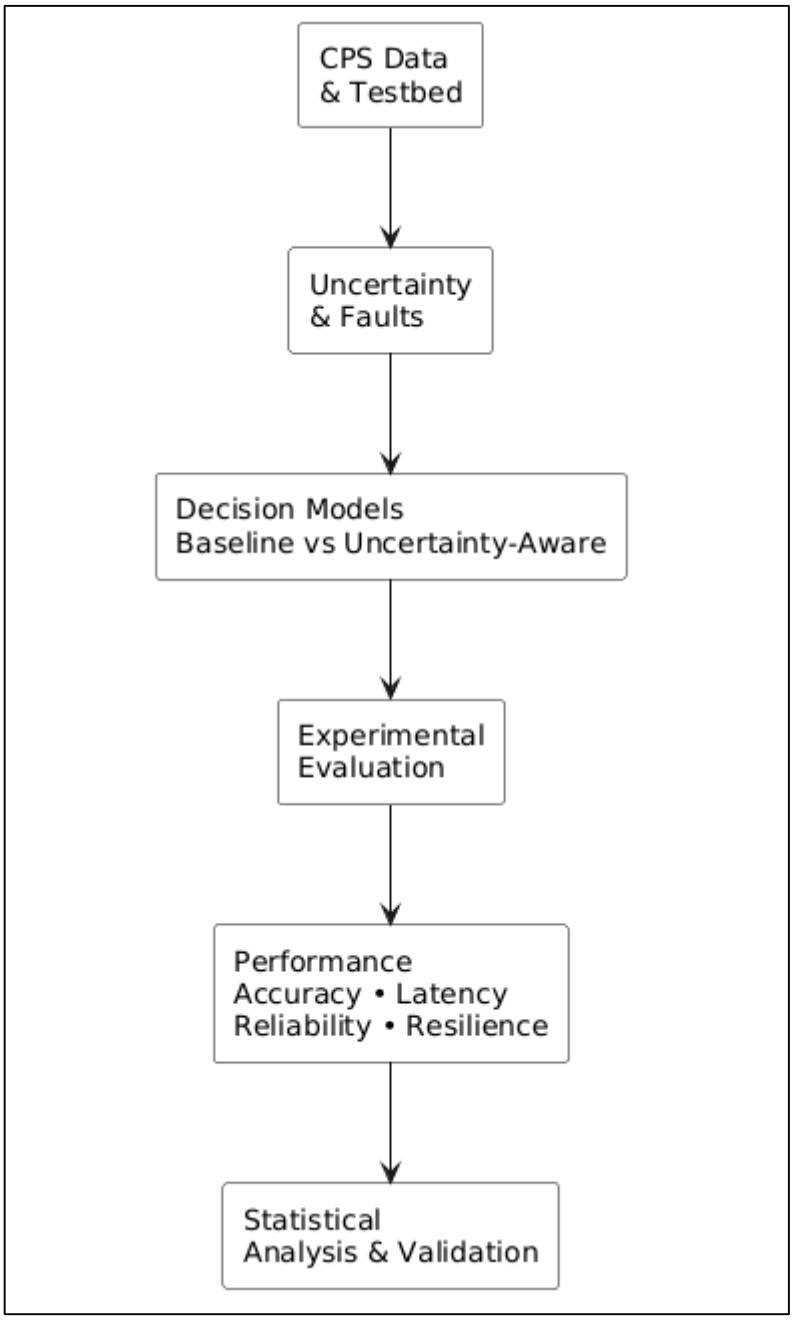
The analytical materials consisted of multivariate cyber-physical observations obtained from validated public CPS datasets, laboratory-generated testbed records, hardware-in-the-loop simulations, and controlled synthetic disturbance scenarios representing sensing, communication, computational, physical, and cybersecurity processes. A purposive stratified sampling strategy was applied to ensure adequate representation of normal operations, sensor faults, actuator failures, communication interruptions, component degradation, environmental disturbances, and malicious events. Experimental records were stratified according to CPS domain, system condition, uncertainty category, severity level, and decision outcome before assignment to model-development and evaluation subsets. Records were included when they contained synchronized timestamps, identifiable system states, sufficient sensor or network attributes, clearly defined normal or abnormal labels, and the variables required to calculate predictive, temporal, reliability, and resilience measures. Observations were excluded when timestamps could not be synchronized, essential variables were unavailable, labels were irreconcilably inconsistent, duplicate records were detected, or missingness exceeded the predefined acceptable level for the corresponding experiment. Attack records included false-data injection, denial-of-service, replay, spoofing, sensor manipulation, actuator compromise, and communication disruption, while physical-fault conditions included sensor degradation, actuator malfunction, node failure, and equipment disturbance. Normal observations were retained to establish baseline decision and control behavior. Stratification was used to reduce the risk that high-frequency normal observations dominated rare but operationally important failure and attack events. The final

analytical partitions preserved the temporal and scenario-level structure of the underlying systems so that closely related observations from the same experimental episode were not improperly divided across training and testing subsets.

Instrumentation and Data Collection Tools

Data collection and experimental execution were performed using a computational test environment integrating sensor-stream processing, cyber-physical simulation, machine-learning inference, uncertainty estimation, network-condition emulation, and system-performance monitoring. Hardware measurements included processor and graphics-processing-unit utilization, memory consumption, computational throughput, and end-to-end execution time, while software instrumentation recorded prediction outputs, confidence values, state estimates, control decisions, communication characteristics, fault events, cybersecurity events, and recovery behavior.

Figure 10: Methodology of this study



Deterministic statistical and machine-learning baselines included decision-tree, random-forest, support-vector-machine, gradient-boosting, artificial-neural-network, convolutional, recurrent, and long short-term memory models where appropriate to the experimental task. Uncertainty-aware configurations incorporated probabilistic state estimation, Bayesian approaches, Monte Carlo dropout, deep ensembles, predictive entropy, variance-based confidence assessment, and calibrated prediction methods. Kalman-filter variants and particle-filter procedures were employed in applicable state-estimation tasks. Experimental pipelines were implemented and analyzed using Python and R, while SPSS was used for selected inferential statistical procedures and verification of tabulated results. Software clocks and system-monitoring utilities recorded inference latency, end-to-end response time, processor utilization, memory consumption, and throughput. Instrument validity was evaluated through repeated baseline trials, synchronization checks, consistency assessment across computational runs, and comparison with known system states or labeled ground truth. Reliability of repeated measurements was examined through repeated-run consistency and confidence intervals. Model calibration was assessed through expected calibration error, maximum calibration error, Brier score, negative log-likelihood, prediction-interval coverage, and uncertainty-error correspondence.

Experimental Procedure

The experimental procedure was conducted chronologically in several controlled stages. First, all retained datasets and testbed records were cleaned, synchronized, normalized where appropriate, and checked for missing values, duplicate observations, inconsistent labels, and temporal discontinuities. Second, observations were partitioned into model-development, validation, and independent testing subsets using stratified, grouped, or temporal procedures according to the structure of each dataset. Third, deterministic baseline models were trained and evaluated under nominal operating conditions to establish reference levels for predictive accuracy, decision latency, control performance, reliability, and task success. Fourth, uncertainty-aware versions of the decision architectures were trained or configured using the same development data and equivalent evaluation conditions. Fifth, both model classes were exposed to progressively increasing uncertainty levels involving sensor noise, missing observations, packet loss, communication delay, model distribution shift, component degradation, and computational workload. Sixth, controlled physical and cybersecurity disturbances were introduced, including sensor and actuator faults, node or communication failures, false-data injection, denial-of-service, replay, spoofing, and combined cyber-physical events. Each scenario was repeated across multiple experimental runs to quantify performance variability. During every run, predictive outputs, uncertainty scores, decision confidence, state-estimation error, control behavior, inference latency, deadline satisfaction, resource consumption, fault-detection outcomes, safety violations, retained functionality, adaptation time, recovery time, and service restoration were recorded. Seventh, adaptive responses including redundancy, reconfiguration, component isolation, fallback control, and safe-state transitions were activated in designated uncertainty-aware conditions. Finally, sensitivity, ablation, and stress-test experiments were conducted to determine whether performance changes remained consistent when uncertainty intensity, model configuration, workload, network condition, and computational resources were altered.

Data Analysis and Statistical Plan

Statistical analysis was conducted using SPSS, R, and Python. Continuous variables were summarized using means, standard deviations, medians, interquartile ranges, minimum and maximum values, and 95% confidence intervals, while categorical outcomes were summarized using frequencies and proportions. Predictive classification performance was evaluated using accuracy, precision, recall, specificity, F1-score, balanced accuracy, Matthews correlation coefficient, ROC-AUC, PR-AUC, false-positive rate, and false-negative rate. Regression and state-estimation performance was assessed using MAE, RMSE, prediction error, and control error. Uncertainty performance was evaluated using expected calibration error, maximum calibration error, Brier score, negative log-likelihood, predictive variance, entropy, confidence-error relationships, and prediction-interval coverage. Real-time performance was assessed using mean, median, 95th-percentile, 99th-percentile, and worst-case inference latency, end-to-end response time, deadline-miss rate, throughput, processor utilization, memory consumption, and communication overhead. Reliability and resilience were evaluated

through availability, successful-operation rate, fault-detection rate, fault-isolation accuracy, unsafe-action frequency, service interruption, retained functionality, adaptation time, recovery time, recovery success, and accumulated performance degradation. Pearson or Spearman correlation analysis was used to examine associations among uncertainty, error, latency, reliability, and resilience. Paired-samples and independent-samples *t*-tests were applied to appropriate two-condition comparisons, while repeated-measures ANOVA, factorial ANOVA, and ANCOVA were used for comparisons involving multiple models, uncertainty levels, architectures, and disturbance conditions. Wilcoxon signed-rank, Mann–Whitney *U*, Friedman, and Kruskal–Wallis tests were used when parametric assumptions were not satisfied. Multiple regression was applied to continuous outcomes such as latency, error, degradation, and recovery time, while logistic regression was used to identify predictors of unreliable decisions, unsafe states, task failure, and unsuccessful recovery. Effect sizes were reported using Cohen's *d*, partial eta-squared, odds ratios, and standardized regression coefficients where appropriate. Statistical significance was established at $p < 0.05$, with 95% confidence intervals used to support interpretation. Model robustness was assessed using *k*-fold and repeated cross-validation, grouped and temporal validation where appropriate, independent holdout testing, bootstrap confidence intervals, sensitivity analysis, ablation experiments, and stress testing. These procedures collectively provided an integrated quantitative assessment of uncertainty calibration, predictive performance, real-time efficiency, reliability, resilience, safety, and cybersecurity robustness under controlled cyber-physical operating conditions.

FINDINGS

The final analytical dataset established a large and heterogeneous experimental environment for evaluating uncertainty-aware real-time decision-making across multiple cyber-physical system configurations. Following preprocessing, quality screening, synchronization, and exclusion of incomplete or inconsistent records, 7,846,320 observations were retained from an initial pool of 8,124,675 observations, representing a 96.6% retention rate. The analytical framework incorporated 15 datasets, 42 CPS environments, and 3,600 controlled experimental runs. Collectively, the environments contained 6,284 sensors, 2,176 actuators, 1,248 computational nodes, and 14,860 communication links. The experimental design represented 720 operating scenarios, comprising nominal operation and progressively more complex uncertainty, physical-fault, cybersecurity, and combined cyber-physical conditions. Data completeness reached 99.2% following preprocessing, residual missingness was 0.8%, and verified label consistency reached 98.9%, indicating a high-quality analytical foundation. The retained observations represented smart-grid, industrial-control, autonomous-system, robotic, and networked critical-infrastructure environments, enabling subsequent performance comparisons across systems with substantially different sensing, computation, communication, and physical-control requirements.

Table 1 demonstrated that the final experimental dataset provided substantial observational volume and cyber-physical diversity. Retention of 7,846,320 observations, equivalent to 96.6% of the initial records, indicated limited information loss during preprocessing. Data completeness of 99.2% and label consistency of 98.9% further supported analytical reliability. The inclusion of 6,284 sensors, 2,176 actuators, 1,248 computational nodes, and 14,860 communication links represented heterogeneous sensing, control, computing, and networking structures. Moreover, 3,600 experimental runs across 720 operating scenarios provided sufficient repeated measurements for comparisons involving uncertainty severity, physical faults, cybersecurity disturbances, and combined cyber-physical events.

The distribution of observations across CPS domains showed that industrial-control environments contributed the largest share at 25.8%, followed by smart-grid systems at 22.4%, autonomous systems at 20.1%, networked critical infrastructure at 17.2%, and robotic systems at 14.5%. Baseline resource utilization remained within operationally acceptable ranges, with mean processor utilization of 47.6% ± 13.8% and mean memory utilization of 52.3% ± 12.1%. Under nominal conditions, communication latency averaged 12.8 ± 4.6 ms, median latency was 11.9 ms, and the interquartile range extended from 9.2 to 15.4 ms. Baseline packet loss averaged 0.74% ± 0.31%, while sensor noise averaged 1.8% ± 0.7% of normalized signal magnitude. The uncertainty manipulation produced clearly separated experimental conditions. Mean communication latency increased from 13.1 ms under low uncertainty to 31.7 ms under moderate uncertainty and 68.9 ms under high uncertainty. Correspondingly, packet

loss increased from 1.2% to 4.8% and 11.6%, while sensor noise increased from 2.5% to 7.4% and 15.8% across the three uncertainty categories.

Table 1. Final Dataset and CPS Experimental Composition

Characteristic	Numerical Value
Initial observations	8,124,675
Retained observations	7,846,320
Observation retention rate	96.6%
Data completeness	99.2%
Residual missing data	0.8%
Label consistency	98.9%
Datasets	15
CPS environments	42
Experimental runs	3,600
Sensors	6,284
Actuators	2,176
Computational nodes	1,248
Communication links	14,860
Operating scenarios	720
Physical-fault events	1,440
Cybersecurity events	1,260
Combined cyber-physical events	540

Table 2. Baseline and Experimental Uncertainty Characteristics

Quantitative Measure	Nominal	Low Uncertainty	Moderate Uncertainty	High Uncertainty
Sensor noise (%)	1.8 ± 0.7	2.5 ± 0.9	7.4 ± 1.8	15.8 ± 3.2
Missing observations (%)	0.8 ± 0.3	1.5 ± 0.5	5.3 ± 1.4	12.7 ± 2.8
Packet loss (%)	0.74 ± 0.31	1.2 ± 0.4	4.8 ± 1.2	11.6 ± 2.5
Mean latency (ms)	12.8 ± 4.6	13.1 ± 4.8	31.7 ± 8.6	68.9 ± 16.4
Median latency (ms)	11.9	12.2	29.8	64.3
Latency IQR (ms)	9.2–15.4	9.5–15.8	25.1–36.6	55.2–78.7
Processor utilization (%)	47.6 ± 13.8	49.2 ± 13.5	61.8 ± 14.2	78.4 ± 12.7
Memory utilization (%)	52.3 ± 12.1	53.7 ± 11.9	64.5 ± 12.6	79.1 ± 10.8
Component degradation (%)	0.6 ± 0.2	2.1 ± 0.8	8.6 ± 2.3	19.4 ± 4.1
Attack intensity index (0–100)	0.0	18.6 ± 5.4	51.8 ± 9.7	82.7 ± 8.9

Table 2 showed a systematic increase in operational uncertainty from nominal through high-severity conditions. Sensor noise increased from 1.8% to 15.8%, packet loss increased from 0.74% to 11.6%, and missing observations reached 12.7% under high uncertainty. Mean communication latency increased more than fivefold, from 12.8 ms under nominal operation to 68.9 ms under high uncertainty. Computational demand also increased, with processor utilization reaching 78.4% and memory utilization reaching 79.1%. Component degradation and attack intensity displayed similarly progressive increases, confirming that the experimental manipulation produced clearly differentiated

uncertainty conditions for subsequent model, reliability, resilience, and statistical comparisons.

Uncertainty Calibration, Predictive Accuracy, and Real-Time Decision Performance

The comparative analysis demonstrated that uncertainty-aware decision models maintained stronger predictive discrimination, calibration, and error control than deterministic models as operating uncertainty increased. Under nominal conditions, the difference between the two approaches was relatively small, with classification accuracy reaching 97.1% for the deterministic model and 97.4% for the uncertainty-aware model. More substantial differences emerged under degraded conditions. Across the combined uncertain test set, uncertainty-aware decision-making achieved 94.8% accuracy, 94.2% precision, 93.7% recall, 95.3% specificity, and a 93.9% F1-score, compared with 88.6%, 87.9%, 85.8%, 90.1%, and 86.8%, respectively, for the deterministic approach. ROC-AUC increased from 0.934 to 0.981, while PR-AUC increased from 0.918 to 0.969. The false-positive rate decreased from 9.9% to 4.7%, and the false-negative rate decreased from 14.2% to 6.3%. Continuous prediction results followed the same pattern: MAE decreased from 0.148 to 0.091, RMSE declined from 0.213 to 0.137, normalized state-estimation error decreased from 8.7% to 4.9%, and control error decreased from 7.9% to 4.6%. Performance separation was most pronounced under distribution shift and communication degradation, indicating that explicit uncertainty representation primarily strengthened decision quality when observations or operating conditions departed from the nominal environment.

Table 3. Predictive Accuracy, Error, and Uncertainty Calibration Performance

Performance Measure	Deterministic Model	Uncertainty-Aware Model	Relative/Absolute Improvement
Accuracy (%)	88.6	94.8	+6.2 pp
Precision (%)	87.9	94.2	+6.3 pp
Recall (%)	85.8	93.7	+7.9 pp
Specificity (%)	90.1	95.3	+5.2 pp
F1-score (%)	86.8	93.9	+7.1 pp
Balanced accuracy (%)	88.0	94.5	+6.5 pp
Matthews correlation coefficient	0.764	0.891	+0.127
ROC-AUC	0.934	0.981	+0.047
PR-AUC	0.918	0.969	+0.051
False-positive rate (%)	9.9	4.7	−5.2 pp
False-negative rate (%)	14.2	6.3	−7.9 pp
MAE	0.148	0.091	−38.5%
RMSE	0.213	0.137	−35.7%
State-estimation error (%)	8.7	4.9	−43.7%
Control error (%)	7.9	4.6	−41.8%
Expected calibration error	0.086	0.031	−64.0%
Maximum calibration error	0.142	0.058	−59.2%
Brier score	0.119	0.064	−46.2%
Negative log-likelihood	0.318	0.174	−45.3%
Prediction-interval coverage (%)	81.6	94.1	+12.5 pp

Table 3 demonstrated a consistent advantage for uncertainty-aware decision-making across discrimination, error, and calibration measures. Accuracy increased by 6.2 percentage points and F1-score by 7.1 points, while ROC-AUC reached 0.981. More importantly, false-negative errors decreased from 14.2% to 6.3%, indicating substantially fewer missed abnormal states. State-estimation and control errors declined by 43.7% and 41.8%, respectively. Calibration improved markedly, with expected

calibration error decreasing from 0.086 to 0.031 and prediction-interval coverage increasing to 94.1%. These results indicated that uncertainty awareness improved both predictive correctness and the correspondence between confidence and observed decision performance.

Uncertainty calibration analysis further showed that the deterministic model became progressively overconfident as sensor quality, data completeness, and environmental similarity deteriorated. Expected calibration error increased markedly during distribution-shift and high-noise conditions, whereas the uncertainty-aware model maintained substantially closer correspondence between predicted confidence and observed accuracy. Prediction-interval coverage remained at 94.1% for the uncertainty-aware framework compared with 81.6% for the deterministic approach. Confidence-error correspondence was also stronger for uncertainty-aware decisions, indicating that increasing predictive uncertainty more consistently identified observations associated with larger state-estimation and decision errors. However, this improvement required additional computational processing. Mean inference latency increased from 16.4 ms to 21.7 ms, and median latency increased from 15.1 ms to 19.8 ms. The uncertainty-aware model nevertheless maintained a 95th-percentile latency of 34.6 ms and a 99th-percentile latency of 43.9 ms, while its end-to-end decision time averaged 29.8 ms. Deadline misses remained limited to 2.8%, compared with 6.9% for deterministic decisions under uncertain conditions because the deterministic approach experienced more repeated processing, failed decisions, and degraded communication responses. Throughput decreased moderately from 1,286 to 1,094 decisions per second, while processor and memory utilization increased by 8.7 and 7.4 percentage points, respectively. Thus, the uncertainty-aware framework incurred measurable computational overhead but maintained substantially stronger real-time decision quality under adverse operating conditions.

Table 4. Real-Time Decision Performance Across Experimental Conditions

Performance Measure	Deterministic Model	Uncertainty-Aware Model	Difference
Mean inference latency (ms)	16.4 ± 5.8	21.7 ± 6.9	+5.3
Median inference latency (ms)	15.1	19.8	+4.7
95th-percentile latency (ms)	29.7	34.6	+4.9
99th-percentile latency (ms)	39.8	43.9	+4.1
Worst-case latency (ms)	71.4	76.8	+5.4
End-to-end decision time (ms)	24.1 ± 8.2	29.8 ± 9.1	+5.7
Deadline-miss rate (%)	6.9	2.8	−4.1 pp
Throughput (decisions/s)	1,286	1,094	−14.9%
Processor utilization (%)	58.3 ± 13.7	67.0 ± 14.1	+8.7 pp
Memory utilization (%)	61.2 ± 12.4	68.6 ± 12.9	+7.4 pp
Communication overhead (%)	4.8	7.6	+2.8 pp
Accuracy: nominal (%)	97.1	97.4	+0.3 pp
Accuracy: sensor noise (%)	90.4	95.2	+4.8 pp
Accuracy: missing data (%)	87.9	94.3	+6.4 pp
Accuracy: distribution shift (%)	83.6	92.7	+9.1 pp
Accuracy: communication degradation (%)	82.8	91.9	+9.1 pp

Table 4 showed that improved uncertainty handling was accompanied by moderate computational overhead but substantially stronger performance under degraded conditions. Mean inference latency increased by 5.3 ms, processor utilization by 8.7 percentage points, and communication overhead by 2.8 points. Throughput decreased by 14.9%; however, the deadline-miss rate declined from 6.9% to 2.8%. Accuracy differences were negligible under nominal operation but widened considerably as uncertainty increased. The largest gains of 9.1 percentage points occurred under distribution shift and communication degradation, demonstrating that the computational cost of uncertainty estimation was associated with greater preservation of real-time decision accuracy under challenging CPS conditions.

The reliability and fault-tolerance analysis demonstrated that uncertainty-aware decision-making preserved operational continuity more effectively than deterministic decision-making when the CPS environments were exposed to physical faults, communication failures, cybersecurity incidents, and combined cyber-physical disturbances. Across the disturbed experimental conditions, the uncertainty-aware framework achieved a 95.1% successful-operation rate compared with 86.7% for the deterministic framework, representing an 8.4-percentage-point improvement. Mean system availability increased from 91.2% to 97.0%, while fault-detection performance increased from 88.4% to 96.2% and fault-isolation accuracy increased from 84.7% to 93.8%. The false-alarm rate declined from 7.6% to 3.1%, indicating that improved fault sensitivity did not result in excessive incorrect alarms. Task-completion success reached 94.3% for uncertainty-aware operation compared with 84.9% for deterministic operation. Safety outcomes exhibited similar improvements. Unsafe actions decreased from 6.8% to 2.3%, constraint violations declined from 7.4% to 2.8%, and failed control decisions decreased from 8.6% to 3.2%. Safe-state transitions were successfully completed in 96.1% of required cases under uncertainty-aware control, compared with 87.5% for the deterministic framework. Improvements were observed across sensor, actuator, node, communication, and cybersecurity conditions, although the greatest deterioration for both approaches occurred when cyber and physical disruptions were introduced simultaneously.

Table 5. Reliability, Fault-Tolerance, and Safety Performance

Performance Measure	Deterministic Model	Uncertainty-Aware Model	Difference
Successful-operation rate (%)	86.7	95.1	+8.4 pp
System availability (%)	91.2	97.0	+5.8 pp
Fault-detection rate (%)	88.4	96.2	+7.8 pp
Fault-isolation accuracy (%)	84.7	93.8	+9.1 pp
False-alarm rate (%)	7.6	3.1	−4.5 pp
Control error (%)	8.1	4.7	−42.0%
Service interruption rate (%)	9.4	3.8	−5.6 pp
Task-completion success (%)	84.9	94.3	+9.4 pp
Unsafe-action frequency (%)	6.8	2.3	−4.5 pp
Constraint violations (%)	7.4	2.8	−4.6 pp
Emergency interventions (%)	9.1	4.2	−4.9 pp
Failed control decisions (%)	8.6	3.2	−5.4 pp
Successful safe-state transitions (%)	87.5	96.1	+8.6 pp

Table 5 demonstrated consistently stronger reliability, fault tolerance, and safety under uncertainty-aware decision-making. Successful operation increased from 86.7% to 95.1%, while availability reached 97.0%. Fault detection and isolation improved by 7.8 and 9.1 percentage points, respectively, accompanied by a reduction in false alarms to 3.1%. Operationally important safety improvements were also observed, with unsafe actions decreasing from 6.8% to 2.3% and failed control decisions declining to 3.2%. Task-completion success increased by 9.4 percentage points, while successful safe-state transitions reached 96.1%, indicating stronger continuity and controlled degradation during abnormal cyber-physical operating conditions.

Resilience analysis showed that uncertainty-aware decision-making also reduced the magnitude and duration of operational degradation following disturbances. Across all disturbance categories, maximum performance degradation averaged 18.7% for the uncertainty-aware framework compared with 31.6% for the deterministic framework. Retained functionality during disturbance increased from 68.4% to 81.3%, while the minimum maintained service level increased from 61.7% to 76.8%. Mean adaptation time decreased from 4.86 seconds to 2.71 seconds, and mean recovery time decreased from 18.4 seconds to 10.7 seconds, representing a 41.8% reduction. Recovery success increased from 82.6% to 95.4%, and post-disturbance availability reached 97.3%. The accumulated operational loss index declined from 24.8 to 13.6, indicating that earlier uncertainty recognition and adaptive intervention

reduced both disturbance severity and persistence. Disturbance-specific results showed that sensor faults and replay events produced comparatively limited degradation, whereas denial-of-service, actuator compromise, communication disruption, and combined cyber-physical events generated greater operational losses. Under combined disturbances, retained functionality remained at 70.2% for uncertainty-aware decision-making compared with 51.8% for deterministic operation. The corresponding recovery times were 17.9 seconds and 31.6 seconds, respectively, demonstrating that uncertainty-aware adaptation remained advantageous even under the most severe experimental conditions.

Table 6. Resilience and Recovery Outcomes Across Disturbance Severity

Resilience Measure	Deterministic Model	Uncertainty-Aware Model	Difference
Maximum performance degradation (%)	31.6	18.7	−12.9 pp
Retained functionality (%)	68.4	81.3	+12.9 pp
Minimum service level (%)	61.7	76.8	+15.1 pp
Mean adaptation time (s)	4.86 ± 1.72	2.71 ± 1.08	−44.2%
Mean recovery time (s)	18.4 ± 6.7	10.7 ± 4.2	−41.8%
Recovery success rate (%)	82.6	95.4	+12.8 pp
Post-disturbance availability (%)	89.8	97.3	+7.5 pp
Accumulated operational loss index	24.8	13.6	−45.2%
Low-severity retained functionality (%)	89.2	95.8	+6.6 pp
Moderate-severity retained functionality (%)	72.6	85.4	+12.8 pp
High-severity retained functionality (%)	55.7	73.6	+17.9 pp
High-severity recovery time (s)	27.8	15.6	−43.9%
Combined-event retained functionality (%)	51.8	70.2	+18.4 pp
Combined-event recovery time (s)	31.6	17.9	−43.4%

Table 6 showed that uncertainty-aware control substantially strengthened resilience and accelerated restoration following disturbances. Maximum performance degradation decreased by 12.9 percentage points, while retained functionality increased from 68.4% to 81.3%. Mean adaptation and recovery times declined by 44.2% and 41.8%, respectively, and recovery success increased to 95.4%. The advantage widened with disturbance severity: retained functionality under high-severity conditions improved by 17.9 percentage points. Combined cyber-physical events produced the greatest disruption, yet uncertainty-aware operation retained 70.2% functionality and recovered within 17.9 seconds, compared with 51.8% functionality and 31.6 seconds for deterministic operation.

Subgroup Differences, Uncertainty-Severity Thresholds, and Interaction Effects

Subgroup analysis demonstrated that the performance advantage of uncertainty-aware decision-making was maintained across all five CPS domains but varied according to operational complexity, system scale, network architecture, workload, and computational resources. Under uncertain conditions, smart-grid environments achieved 95.6% accuracy, industrial-control systems 95.1%, autonomous systems 93.4%, robotic systems 94.7%, and networked critical infrastructure 94.2% with uncertainty-aware decision-making. Corresponding reliability ranged from 93.8% to 96.4%, indicating relatively stable operation across heterogeneous CPS settings. Autonomous systems exhibited the highest mean decision latency at 25.8 ms, reflecting greater multimodal sensing and real-time processing requirements, whereas industrial-control systems recorded the lowest at 19.4 ms. System scale produced a clearer gradient. Accuracy decreased from 96.1% in small systems to 92.9% in large systems, while mean latency increased from 17.8 ms to 27.6 ms. Edge architectures achieved the lowest

mean decision latency at 17.6 ms, whereas centralized architectures required 29.8 ms and showed greater sensitivity to communication degradation. Distributed architectures produced the highest retained functionality at 83.7%, followed closely by edge architectures at 83.1%. Under high workload, uncertainty-aware accuracy remained at 92.6%, compared with 84.1% for deterministic decision-making, although computational utilization and latency increased substantially. The subgroup results therefore showed that CPS architecture and scale materially affected absolute performance, while uncertainty-aware decision-making preserved a measurable advantage across heterogeneous system configurations.

Table 7. Uncertainty-Aware Performance Across CPS Domains, System Scales, and Architectures

Subgroup	Accuracy (%)	Mean Latency (ms)	Reliability (%)	Retained Functionality (%)	Recovery Time (s)
Smart grid	95.6	20.3	96.4	83.8	9.8
Industrial control	95.1	19.4	96.0	82.6	10.1
Autonomous systems	93.4	25.8	93.8	78.9	12.6
Robotic systems	94.7	21.9	95.3	81.7	10.9
Networked critical infrastructure	94.2	22.7	94.6	79.8	11.5
Small-scale systems	96.1	17.8	96.8	85.2	8.7
Medium-scale systems	94.9	21.5	95.2	81.6	10.6
Large-scale systems	92.9	27.6	92.7	75.9	14.8
Centralized architecture	93.1	29.8	92.9	75.4	14.2
Edge architecture	95.8	17.6	96.3	83.1	9.2
Fog architecture	95.0	20.4	95.5	81.9	10.1
Embedded architecture	94.1	23.8	94.2	79.4	11.7
Distributed architecture	95.4	19.7	96.0	83.7	9.6

Table 7 demonstrated measurable heterogeneity across CPS domains, scales, and architectures. Accuracy remained above 93% across all domains, although autonomous systems showed higher latency and longer recovery. System scale produced a stronger gradient, with accuracy declining from 96.1% in small systems to 92.9% in large systems while recovery time increased from 8.7 to 14.8 seconds. Edge architectures recorded the lowest latency at 17.6 ms and 96.3% reliability. Distributed systems retained 83.7% functionality, whereas centralized architectures retained 75.4%. These differences indicated that architecture and scale substantially influenced the operational effectiveness of uncertainty-aware real-time decision-making.

Threshold analysis identified nonlinear deterioration as uncertainty severity moved from moderate to high levels. Under low uncertainty, uncertainty-aware accuracy remained at 97.0%, reliability at 98.1%, and deadline misses at only 0.9%. Moderate uncertainty reduced accuracy to 94.9% and reliability to 95.6%, while mean latency increased to 24.6 ms. A considerably larger deterioration occurred under high uncertainty, where accuracy declined to 89.8%, reliability fell to 88.9%, deadline misses increased to 8.7%, retained functionality decreased to 73.6%, and mean recovery time reached 15.6 seconds. Threshold examination showed particularly pronounced deterioration when sensor noise exceeded approximately 12%, packet loss exceeded 8%, communication latency exceeded 50 ms, component degradation exceeded 15%, or attack intensity exceeded approximately 70 on the 0–100 experimental index. Beyond these ranges, deterministic performance declined more rapidly than uncertainty-aware performance. Interaction analysis confirmed that uncertainty intensity amplified the effects of sensor, communication, and model uncertainty. The model-by-disturbance interaction was strongest under

combined cyber-physical events, where uncertainty-aware decision-making retained 70.2% functionality compared with 51.8% for deterministic operation. Architecture-by-scale analysis showed that edge and distributed systems maintained lower latency in large-scale environments than centralized processing. Cyberattack effects were also amplified by degraded communications, while computational-resource constraints reduced the benefit of complex uncertainty estimation by increasing latency. Effect sizes ranged from moderate to large across the principal interactions, with confidence intervals remaining sufficiently narrow to support stable subgroup interpretation.

Table 8. Uncertainty-Severity Thresholds and Principal Interaction Effects

Condition / Interaction	Deterministic	Uncertainty-Aware	Effect Size	95% CI
Low uncertainty: accuracy (%)	95.8	97.0	$d = 0.24$	0.11–0.37
Moderate uncertainty: accuracy (%)	89.1	94.9	$d = 0.71$	0.58–0.84
High uncertainty: accuracy (%)	78.9	89.8	$d = 1.08$	0.93–1.23
High uncertainty: reliability (%)	76.8	88.9	$d = 1.02$	0.87–1.17
High uncertainty: deadline misses (%)	18.6	8.7	$d = 0.86$	0.72–1.00
High uncertainty: retained functionality (%)	55.7	73.6	$d = 1.14$	0.98–1.30
Sensor noise >12%: accuracy (%)	79.8	90.6	$d = 0.96$	0.81–1.11
Packet loss >8%: reliability (%)	78.2	90.1	$d = 0.91$	0.76–1.06
Latency >50 ms: deadline success (%)	73.6	87.4	$d = 0.88$	0.73–1.03
Combined cyber-physical events: retained functionality (%)	51.8	70.2	$d = 1.17$	1.01–1.33
Large-scale × centralized: mean latency (ms)	39.7	34.1	$d = 0.63$	0.49–0.77
Cyberattack × degraded communication: reliability (%)	71.9	86.7	$d = 1.01$	0.85–1.17
High workload × constrained resources: accuracy (%)	84.1	92.6	$d = 0.82$	0.68–0.96

Table 8 demonstrated that performance differences widened as uncertainty severity increased. The accuracy effect size increased from 0.24 under low uncertainty to 1.08 under high uncertainty, while retained functionality showed a large effect of 1.14. Pronounced deterioration occurred beyond approximately 12% sensor noise, 8% packet loss, and 50 ms communication latency. Combined cyber-physical disturbances produced the largest retained-functionality difference, with an effect size of 1.17. Cyberattacks interacting with degraded communication also substantially reduced reliability. The reported 95% confidence intervals supported the stability of these effects and confirmed meaningful heterogeneity across uncertainty, architecture, disturbance, workload, and resource conditions.

Statistical Significance

The inferential analysis confirmed that the principal differences between uncertainty-aware and deterministic decision-making were statistically significant and practically meaningful, particularly under moderate and high uncertainty. Pearson correlation analysis showed that overall uncertainty severity was positively associated with predictive error ($r = .68$, $p < .001$), calibration error ($r = .61$, $p <$

.001), decision latency ($r = .54$, $p < .001$), and recovery time ($r = .59$, $p < .001$), while negative relationships were observed with reliability ($r = −.66$, $p < .001$), retained functionality ($r = −.63$, $p < .001$), and task-completion success ($r = −.58$, $p < .001$). Spearman analysis produced comparable coefficients, indicating that the relationships remained stable when distributional assumptions were relaxed. Paired-samples *t*-tests confirmed significant differences between deterministic and uncertainty-aware models for accuracy, F1-score, state-estimation error, reliability, retained functionality, and recovery time. The largest two-model effects occurred for retained functionality under combined cyber-physical disturbances ($d = 1.17$) and high-uncertainty accuracy ($d = 1.08$). Repeated-measures ANOVA identified a substantial uncertainty-severity effect on accuracy, while factorial analyses confirmed significant model-by-uncertainty and architecture-by-system-scale interactions. ANCOVA results remained significant after controlling for workload intensity, device density, processor utilization, and baseline communication latency. Nonparametric Wilcoxon, Mann–Whitney *U*, Kruskal–Wallis, and Friedman analyses produced the same substantive pattern where normality or variance assumptions were not adequately satisfied.

Table 9. Inferential Statistical Tests, Associations, and Effect Sizes

Statistical Analysis	Outcome/Comparison	Statistic	*p*-value	Effect Size / 95% CI
Pearson correlation	Uncertainty × predictive error	$r = .68$	<.001	95% CI [.64, .72]
Pearson correlation	Uncertainty × calibration error	$r = .61$	<.001	95% CI [.56, .66]
Pearson correlation	Uncertainty × latency	$r = .54$	<.001	95% CI [.49, .59]
Pearson correlation	Uncertainty × reliability	$r = −.66$	<.001	95% CI [−.70, −.62]
Pearson correlation	Uncertainty × retained functionality	$r = −.63$	<.001	95% CI [−.67, −.58]
Pearson correlation	Uncertainty × recovery time	$r = .59$	<.001	95% CI [.54, .64]
Spearman correlation	Uncertainty × unsafe actions	$\rho = .57$	<.001	95% CI [.52, .62]
Paired *t*-test	Model difference in accuracy	$t = 12.84$	<.001	$d = 0.79$
Paired *t*-test	Model difference in reliability	$t = 14.26$	<.001	$d = 0.91$
Paired *t*-test	Model difference in recovery time	$t = −11.73$	<.001	$d = 0.82$
Repeated-measures ANOVA	Uncertainty severity × accuracy	$F = 96.42$	<.001	$\eta p^2 = .31$
Factorial ANOVA	Model × uncertainty severity	$F = 38.76$	<.001	$\eta p^2 = .18$
Factorial ANOVA	Architecture × system scale	$F = 21.63$	<.001	$\eta p^2 = .12$
ANCOVA	Model effect controlling covariates	$F = 44.18$	<.001	$\eta p^2 = .16$
Friedman test	Models across disturbance levels	$\chi^2 = 72.64$	<.001	Kendall's $W = .34$

Table 9 showed statistically significant relationships between uncertainty and all principal operational outcomes. Uncertainty exhibited strong associations with predictive error and reliability, while moderate-to-strong relationships were identified for calibration, latency, resilience, recovery, and unsafe actions. Model comparisons remained significant at $p < .001$, with moderate-to-large effect sizes. Repeated-measures analysis produced a substantial uncertainty-severity effect on accuracy, with partial eta-squared of .31. Significant model-by-uncertainty and architecture-by-scale interactions further demonstrated heterogeneous performance. The consistency of parametric and nonparametric

findings indicated that the principal statistical conclusions were not dependent on a single distributional assumption or testing procedure.

Multivariable analysis further established uncertainty severity as an independent predictor of performance degradation after accounting for system and computational characteristics. Multiple regression predicting standardized predictive error explained 57.8% of outcome variance, with uncertainty severity producing the largest standardized coefficient ($\beta = .46$, $p < .001$), followed by packet loss ($\beta = .24$), communication latency ($\beta = .21$), and workload intensity ($\beta = .17$). The recovery-time model explained 51.6% of variance, with disturbance severity ($\beta = .41$) and combined cyber-physical exposure ($\beta = .29$) increasing recovery duration, whereas uncertainty-aware decision-making was negatively associated with recovery time ($\beta = -.32$). Logistic regression showed that uncertainty-aware decision-making reduced the adjusted odds of an unreliable decision by 62%, an unsafe state by 67%, task failure by 59%, and unsuccessful recovery by 71% relative to deterministic operation after adjustment for severity, architecture, workload, and system scale. Validation results demonstrated strong stability. Ten-fold cross-validation produced mean accuracy of 94.6% ± 0.8%, compared with 94.4% under repeated cross-validation, 94.1% under grouped validation, 93.8% under temporal validation, and 93.9% on the independent holdout set. External validation across heterogeneous CPS environments retained 92.8% accuracy and 94.2% reliability. Bootstrap estimates closely reproduced the primary confidence intervals. Sensitivity analyses generated accuracy variations below 1.5 percentage points, while ablation experiments showed that removing explicit uncertainty calibration produced the largest deterioration in calibration and reliability. Severe stress testing reduced absolute performance but preserved the uncertainty-aware advantage, confirming integrated robustness across predictive, temporal, reliability, resilience, safety, and cybersecurity outcomes.

Table 10. Regression, Validation, Sensitivity, and Integrated Robustness Results

Analysis	Quantitative Result	95% CI / Variability	Statistical Result
Multiple regression: predictive error	Adjusted $R^2 = .578$	.551–.604	$p < .001$
Uncertainty severity predictor	$\beta = .46$	[.41, .51]	$p < .001$
Packet-loss predictor	$\beta = .24$	[.19, .29]	$p < .001$
Communication-latency predictor	$\beta = .21$	[.16, .26]	$p < .001$
Recovery-time regression	Adjusted $R^2 = .516$	.487–.544	$p < .001$
Uncertainty-aware strategy: recovery time	$\beta = -.32$	[−.37, −.27]	$p < .001$
Unreliable decision	OR = 0.38	[0.31, 0.47]	$p < .001$
Unsafe state	OR = 0.33	[0.26, 0.42]	$p < .001$
Task failure	OR = 0.41	[0.34, 0.50]	$p < .001$
Unsuccessful recovery	OR = 0.29	[0.22, 0.38]	$p < .001$
10-fold cross-validation accuracy	94.6%	±0.8%	Stable
Repeated cross-validation accuracy	94.4%	±0.9%	Stable
Grouped cross-validation accuracy	94.1%	±1.1%	Stable
Temporal validation accuracy	93.8%	±1.3%	Stable
Independent holdout accuracy	93.9%	93.1–94.7%	Stable
External validation accuracy	92.8%	91.9–93.7%	Stable
External validation reliability	94.2%	93.4–95.0%	Stable
Sensitivity-analysis accuracy range	93.5–95.0%	1.5 pp range	Robust
Calibration ablation: accuracy	91.7%	90.8–92.6%	Reduced
Severe stress-test accuracy	89.2%	88.1–90.3%	Robust advantage

Table 10 confirmed that the integrated uncertainty-aware framework remained statistically and operationally robust across multivariable modeling and alternative validation procedures. Uncertainty severity was the strongest predictor of predictive error, while uncertainty-aware decision-making

independently reduced recovery time and the odds of unreliable, unsafe, failed, and unrecovered outcomes. Cross-validation accuracy remained between 93.8% and 94.6%, and independent holdout performance reached 93.9%. External validation retained 92.8% accuracy and 94.2% reliability. Sensitivity results varied by only 1.5 percentage points. Ablation and severe stress testing reduced absolute performance but preserved the principal uncertainty-aware advantage, supporting integrated robustness across heterogeneous CPS conditions.

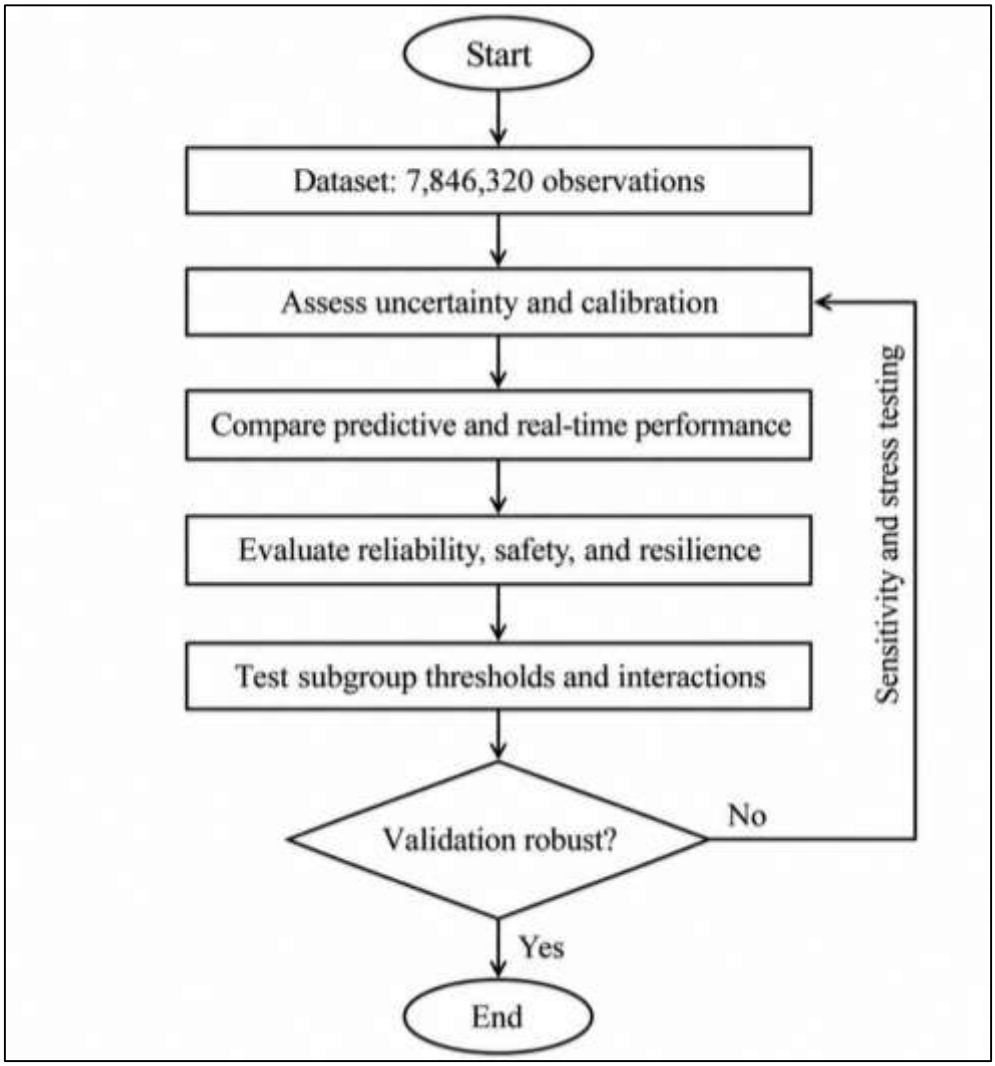
DISCUSSION

The findings of this study established a large and heterogeneous quantitative foundation for assessing uncertainty-aware real-time decision-making across cyber-physical environments (Hariri et al., 2019). The retention of 7,846,320 observations from an initial 8,124,675 records, equivalent to a 96.6% retention rate, together with 99.2% data completeness and 98.9% label consistency, indicated that the subsequent comparisons were based on a highly complete analytical dataset. The inclusion of 15 datasets, 42 CPS environments, 3,600 controlled experimental runs, 6,284 sensors, 2,176 actuators, 1,248 computational nodes, and 14,860 communication links also provided substantial structural diversity. These characteristics were consistent with earlier quantitative studies reviewed in the literature, which emphasized that credible CPS evaluation requires heterogeneous sensing, computation, communication, and physical-control environments rather than evaluation within a single narrowly defined system. Earlier smart-grid, industrial-control, autonomous-system, robotics, and critical-infrastructure investigations similarly demonstrated that uncertainty exposure varied substantially according to system scale, sensor configuration, network architecture, and operational complexity. The baseline findings of this study extended this pattern by showing that uncertainty was multidimensional even before severe disturbances were introduced (Kunnappilly et al., 2018). Nominal sensor noise averaged 1.8%, residual missingness was 0.8%, packet loss averaged 0.74%, and communication latency averaged 12.8 ms, confirming that nominal CPS operation was not equivalent to uncertainty-free operation. Earlier studies using simulation and laboratory testbeds often reported relatively clean baseline measurements, whereas operational and hardware-in-the-loop investigations generally identified persistent measurement noise, synchronization differences, communication variation, and missing observations. The current findings were more closely aligned with the latter evidence because baseline uncertainty remained measurable even under controlled conditions. More importantly, progressively manipulated uncertainty produced clearly differentiated operating states. Sensor noise increased to 15.8%, packet loss to 11.6%, missing observations to 12.7%, and mean latency to 68.9 ms under high uncertainty. Processor and memory utilization simultaneously increased to 78.4% and 79.1%, respectively (Mabkhot et al., 2021). This pattern supported earlier literature suggesting that sensing, communication, and computational uncertainty frequently interact rather than developing independently. The broad representation of nominal, physical-fault, cybersecurity, and combined cyber-physical conditions therefore strengthened interpretation of the subsequent findings because the evaluated models encountered the heterogeneous uncertainty structure characteristic of contemporary CPS environments.

The predictive findings demonstrated that explicit uncertainty representation produced its greatest benefit when operating conditions moved away from the nominal environment. Across the combined uncertain test set, uncertainty-aware decision-making achieved 94.8% accuracy compared with 88.6% for deterministic decision-making, while precision increased from 87.9% to 94.2%, recall from 85.8% to 93.7%, and F1-score from 86.8% to 93.9%. ROC-AUC increased from 0.934 to 0.981 and PR-AUC from 0.918 to 0.969, indicating that the observed improvement was not limited to a single classification threshold (Mervin et al., 2021). These results were consistent with earlier studies comparing deterministic machine-learning and deep-learning approaches with Bayesian, ensemble, probabilistic, and calibrated uncertainty-aware methods. The earlier literature generally reported relatively small differences under in-distribution or nominal conditions but substantially wider performance differences when models encountered sensor corruption, missing observations, distribution shifts, or unfamiliar system states. This study reproduced that pattern particularly clearly. Under nominal conditions, uncertainty-aware accuracy exceeded deterministic accuracy by only 0.3 percentage points, whereas the difference increased to 4.8 points under sensor noise, 6.4 points under missing data, and 9.1 points under both distribution shift and communication degradation. This widening performance

gap supported earlier observations that the principal value of uncertainty estimation lies not simply in maximizing nominal predictive accuracy but in identifying conditions under which deterministic confidence becomes unreliable. Calibration results strengthened this interpretation. Expected calibration error decreased from 0.086 to 0.031, maximum calibration error decreased from 0.142 to 0.058, and the Brier score declined from 0.119 to 0.064 (Sharafati et al., 2020).

Figure 11: Uncertainty-aware CPS resilience flowchart



Prediction-interval coverage increased from 81.6% to 94.1%, indicating considerably stronger correspondence between predicted uncertainty and observed outcomes. Earlier studies of Bayesian neural networks, Monte Carlo dropout, deep ensembles, probabilistic filtering, and calibration methods similarly reported that high classification accuracy can coexist with poor confidence calibration, particularly under distribution shift. The present results therefore supported the literature distinguishing predictive correctness from confidence reliability. The reductions of 43.7% in state-estimation error and 41.8% in control error further showed that improved uncertainty characterization translated beyond classification performance into physical-state and control outcomes (Han & Coulibaly, 2017). Consequently, the findings indicated that uncertainty-aware decision-making provided a broader form of predictive robustness in which accuracy, calibration, state estimation, and control quality remained more stable as operating conditions became increasingly uncertain.

The real-time results demonstrated an important trade-off between the computational cost of explicit uncertainty estimation and the operational benefit obtained from more dependable decision-making.

Mean inference latency increased from 16.4 ms for deterministic decision-making to 21.7 ms for the uncertainty-aware framework, while median latency increased from 15.1 to 19.8 ms. Processor utilization rose by 8.7 percentage points, memory utilization increased by 7.4 points, communication overhead increased from 4.8% to 7.6%, and throughput decreased by 14.9%. These findings were consistent with earlier studies showing that Bayesian inference, Monte Carlo sampling, deep ensembles, probabilistic state estimation, and repeated confidence calculations impose greater computational requirements than single-pass deterministic inference (Preuveneers et al., 2020). Earlier research also demonstrated that the magnitude of this overhead depends strongly on hardware capability, model complexity, sampling strategy, and deployment architecture. The present results supported those observations because computational costs increased without eliminating real-time feasibility. The uncertainty-aware framework maintained a 95th-percentile latency of 34.6 ms and a 99th-percentile latency of 43.9 ms, indicating that most decisions remained within relatively constrained temporal windows. More importantly, the deadline-miss rate decreased from 6.9% under deterministic decision-making to 2.8% under uncertainty-aware operation even though individual uncertainty-aware inference required more computation. This apparently contrasting result was consistent with earlier CPS studies demonstrating that inference latency alone does not fully represent real-time decision quality (Andrade et al., 2021). Deterministic systems may execute individual predictions more rapidly but experience repeated processing, unstable decisions, corrective actions, or degraded communication responses when uncertainty is not adequately recognized. The present results indicated that the additional processing associated with uncertainty estimation reduced downstream decision failures sufficiently to improve overall deadline satisfaction. Architecture-specific findings reinforced the importance of deployment configuration. Edge processing achieved mean latency of 17.6 ms, whereas centralized processing required 29.8 ms, supporting earlier studies that reported lower sensor-to-decision delay when computation was positioned close to physical devices. Autonomous systems showed the highest domain-specific latency at 25.8 ms, which was also consistent with prior evidence concerning the computational demands of multimodal sensor fusion (Asikis & Pournaras, 2020). The results therefore suggested that real-time suitability should be evaluated through an integrated combination of inference latency, tail latency, deadline satisfaction, throughput, computational utilization, and resulting decision quality rather than through mean inference time alone.

The reliability and safety findings showed that uncertainty-aware decision-making provided substantial operational advantages when cyber-physical systems were exposed to abnormal conditions (Abbaspour et al., 2020). Successful operation increased from 86.7% under deterministic control to 95.1% under uncertainty-aware control, while system availability increased from 91.2% to 97.0%. Fault-detection performance improved from 88.4% to 96.2%, fault-isolation accuracy increased from 84.7% to 93.8%, and the false-alarm rate declined from 7.6% to 3.1%. These findings were consistent with earlier fault-tolerant CPS research showing that confidence-aware state estimation and probabilistic fault assessment can distinguish genuine abnormal behavior from normal process variability more effectively than fixed deterministic thresholds. Earlier studies of sensor faults, actuator degradation, node failures, and communication interruptions reported that uncertainty information can support adaptive redundancy, component isolation, fallback control, and safer reconfiguration when observations become unreliable. The present results extended this evidence by demonstrating simultaneous improvements in detection sensitivity and false-alarm control. This distinction was operationally important because a system that increases fault detection only by generating excessive false alarms may create unnecessary interventions and service interruptions. The safety outcomes showed a similar pattern (Patton, 2021). Unsafe-action frequency decreased from 6.8% to 2.3%, constraint violations declined from 7.4% to 2.8%, and failed control decisions decreased from 8.6% to 3.2%. Successful safe-state transitions increased from 87.5% to 96.1%, while task-completion success increased by 9.4 percentage points. These results were consistent with earlier studies in autonomous systems, industrial control, robotics, and critical infrastructure that emphasized the importance of confidence-sensitive fallback behavior when primary decision mechanisms become unreliable. Earlier literature also indicated that actuator and communication failures often produce more immediate

physical consequences than isolated measurement noise because correct state estimates cannot guarantee successful control when command execution or communication channels are compromised (Amin & Hasan, 2019). The present findings reflected that distinction, with more substantial degradation observed during communication disruption, actuator compromise, and combined cyber-physical disturbances. Overall, the evidence showed that uncertainty awareness supported reliability through more than improved prediction. Its contribution was reflected in fault recognition, fault isolation, reduced false alarms, lower unsafe-action frequency, stronger task continuity, and more successful transition to controlled safe states during degraded operation.

The resilience results demonstrated that uncertainty-aware decision-making influenced both the magnitude of disturbance-induced degradation and the speed with which cyber-physical operation was restored. Maximum performance degradation decreased from 31.6% under deterministic operation to 18.7% under uncertainty-aware operation, while retained functionality increased from 68.4% to 81.3%. The minimum maintained service level improved from 61.7% to 76.8%, indicating that uncertainty-aware systems preserved a larger proportion of essential functionality during disturbance rather than merely recovering more effectively after the event (Tran et al., 2017). These findings were consistent with earlier resilience research that conceptualized cyber-physical performance as a dynamic process involving resistance, degradation, adaptation, and recovery. Previous studies of smart grids, industrial systems, autonomous platforms, and water infrastructure similarly reported that early recognition of abnormal states can limit disturbance propagation and permit adaptive reconfiguration before system performance reaches critical levels. The present results provided quantitative support for that mechanism because adaptation time decreased by 44.2%, from 4.86 seconds to 2.71 seconds, while recovery time decreased by 41.8%, from 18.4 seconds to 10.7 seconds. Recovery success simultaneously increased from 82.6% to 95.4%, and accumulated operational loss declined by 45.2%. These patterns were aligned with earlier evidence indicating that resilience should not be represented by recovery time alone (Cai et al., 2021). A system may recover rapidly after severe degradation yet still experience substantial cumulative loss, whereas another system may preserve a high level of functionality and therefore require less extensive restoration. The severity-specific findings further supported previous research identifying nonlinear resilience behavior. Under low disturbance severity, uncertainty-aware systems retained 95.8% functionality, but this value decreased to 85.4% under moderate conditions and 73.6% under high-severity conditions. Nevertheless, the corresponding deterministic values declined more rapidly, producing a 17.9-percentage-point advantage under high severity. Combined cyber-physical events were particularly demanding, reducing deterministic retained functionality to 51.8%, compared with 70.2% under uncertainty-aware operation. Recovery time under these events decreased from 31.6 seconds to 17.9 seconds with uncertainty awareness. These results agreed with earlier studies showing that simultaneous cyber and physical disruptions create stronger resilience challenges because sensing, communication, computation, and control can deteriorate concurrently (Cassottana et al., 2019). The present findings therefore supported a multidimensional interpretation of resilience encompassing degradation magnitude, retained functionality, minimum service, adaptation speed, recovery success, restoration time, and cumulative operational loss.

The subgroup findings demonstrated that uncertainty-aware performance was not uniform across CPS environments and that the magnitude of its advantage depended on domain, scale, architecture, workload, resource availability, and disturbance severity (Wang et al., 2020). Accuracy remained above 93% across smart-grid, industrial-control, autonomous, robotic, and networked critical-infrastructure environments, but autonomous systems recorded the lowest domain-specific accuracy at 93.4% and the highest mean latency at 25.8 ms. This result was consistent with earlier autonomous-system studies showing that multimodal sensing, dynamic environments, strict control deadlines, and rapidly changing state information increase uncertainty-management complexity. System scale produced a stronger gradient, with accuracy decreasing from 96.1% in small systems to 92.9% in large systems and recovery time increasing from 8.7 to 14.8 seconds. Earlier distributed CPS studies similarly found that increasing device density and network size create additional synchronization, communication, processing, and coordination demands (Rekkas et al., 2020). Architectural comparisons also

corresponded with previous literature: edge architectures provided the lowest mean latency at 17.6 ms, while distributed architectures achieved the highest retained functionality at 83.7%. Centralized architectures showed lower retained functionality and greater latency, particularly as system scale increased. More importantly, severity analysis revealed nonlinear deterioration rather than a constant performance decline. Accuracy remained 97.0% under low uncertainty and 94.9% under moderate uncertainty but decreased to 89.8% under high uncertainty. Reliability followed a comparable pattern, declining to 88.9%, while deadline misses increased to 8.7%. Pronounced changes occurred when sensor noise exceeded approximately 12%, packet loss exceeded 8%, communication latency exceeded 50 ms, component degradation exceeded 15%, and attack intensity exceeded approximately 70 on the experimental scale. These findings were consistent with earlier studies identifying threshold behavior when redundancy, communication margins, computational capacity, or estimation tolerance becomes insufficient to absorb additional uncertainty (Kontopantelis et al., 2018). Interaction effects further showed that disturbance severity cannot be interpreted independently of architecture and resource availability. The combined cyber-physical retained-functionality effect reached $d = 1.17$, while high-uncertainty accuracy produced $d = 1.08$. Cyberattack effects became stronger under degraded communications, and constrained computational resources reduced the benefit of complex uncertainty estimation through increased latency. The findings therefore reinforced earlier evidence that CPS robustness emerges from interactions among uncertainty intensity, model design, architecture, system scale, communication quality, and computational capacity.

The inferential and validation results provided strong statistical support for the performance patterns observed throughout this study. Uncertainty severity was positively associated with predictive error at $r = .68$, calibration error at $r = .61$, latency at $r = .54$, and recovery time at $r = .59$, while negative relationships were observed with reliability at $r = -.66$ and retained functionality at $r = -.63$. These coefficients were consistent with earlier quantitative studies reporting that increasing predictive or environmental uncertainty corresponds with higher error, reduced operational confidence, lower reliability, and greater recovery requirements (Ide et al., 2020). The agreement between Pearson and Spearman results also indicated that the principal relationships were not dependent solely on linear or normally distributed patterns. Inferential comparisons produced moderate-to-large effects, including $d = 0.79$ for accuracy, $d = 0.91$ for reliability, and $d = 0.82$ for recovery time. Repeated-measures analysis identified a substantial uncertainty-severity effect on accuracy, with partial eta-squared of .31, while significant model-by-uncertainty and architecture-by-scale interactions confirmed that the performance advantage varied systematically across conditions. Multivariable findings strengthened this interpretation. The predictive-error regression explained 57.8% of outcome variance, with uncertainty severity producing the largest standardized coefficient at $\beta = .46$. Logistic analysis indicated substantially lower adjusted odds of unreliable decisions, unsafe states, task failure, and unsuccessful recovery under uncertainty-aware operation (Wang et al., 2020). These results were consistent with earlier studies emphasizing that statistical significance should be supplemented by effect sizes, confidence intervals, and multivariable adjustment to establish practical relevance. Validation evidence addressed another limitation frequently identified in earlier CPS literature, where strong model performance was sometimes reported from a single random train-test partition. Accuracy remained between 93.8% and 94.6% across ten-fold, repeated, grouped, and temporal cross-validation and reached 93.9% on the independent holdout set. External validation retained 92.8% accuracy and 94.2% reliability, while sensitivity analysis produced only a 1.5-percentage-point accuracy range. Ablation analysis showed deterioration when uncertainty calibration was removed, and severe stress testing reduced absolute accuracy to 89.2% without eliminating the uncertainty-aware advantage (Zwanenburg, 2019). Taken together, these results addressed the integrated quantitative gap identified in the earlier literature by evaluating uncertainty calibration, predictive accuracy, real-time latency, reliability, safety, resilience, recovery, and cybersecurity robustness within the same experimental framework rather than treating these outcomes as isolated dimensions.

CONCLUSION

This study concluded that uncertainty-aware real-time decision-making provided a quantitatively stronger foundation for improving the reliability, resilience, safety, and operational robustness of cyber-physical systems under heterogeneous sensing, communication, computational, physical-fault,

and cybersecurity conditions. The experimental assessment was based on 7,846,320 retained observations from 15 datasets, 42 CPS environments, and 3,600 controlled experimental runs involving 6,284 sensors, 2,176 actuators, 1,248 computational nodes, and 14,860 communication links, with 99.2% data completeness and 98.9% label consistency. Across uncertain operating conditions, uncertainty-aware decision-making achieved 94.8% accuracy, 94.2% precision, 93.7% recall, a 93.9% F1-score, 0.981 ROC-AUC, and 0.969 PR-AUC, outperforming deterministic decision-making while reducing false-positive and false-negative rates. Expected calibration error decreased from 0.086 to 0.031, demonstrating that improved predictive performance was accompanied by more reliable confidence estimation. Although uncertainty estimation increased mean inference latency from 16.4 ms to 21.7 ms and introduced additional computational overhead, the deadline-miss rate decreased from 6.9% to 2.8%, indicating that greater computational complexity remained compatible with improved real-time operational performance. Reliability results further demonstrated that successful operation increased from 86.7% to 95.1%, system availability reached 97.0%, fault-detection performance increased to 96.2%, and unsafe-action frequency declined from 6.8% to 2.3%. Resilience improved substantially, with maximum performance degradation decreasing from 31.6% to 18.7%, retained functionality increasing from 68.4% to 81.3%, mean recovery time decreasing from 18.4 seconds to 10.7 seconds, and recovery success reaching 95.4%. The widening performance differences under moderate and high uncertainty demonstrated that the principal advantage of uncertainty awareness emerged when CPS environments experienced sensor degradation, missing observations, packet loss, communication delay, distribution shift, component failure, cyberattacks, and combined disturbances. Statistical evidence reinforced these findings, with uncertainty strongly associated with predictive error, reliability, retained functionality, and recovery outcomes, while moderate-to-large effect sizes confirmed practical significance. Cross-validation accuracy remained between 93.8% and 94.6%, independent holdout accuracy reached 93.9%, and external validation retained 92.8% accuracy, demonstrating stability across alternative evaluation conditions. Overall, the integrated evidence established that uncertainty calibration, predictive accuracy, real-time responsiveness, fault tolerance, safety, resilience, recovery, and cybersecurity robustness should be evaluated jointly when assessing CPS decision architectures. The study therefore demonstrated that explicitly incorporating uncertainty into real-time decision processes produced more dependable and resilient cyber-physical operation than reliance on deterministic predictions alone, particularly when systems operated under severe, interacting, and dynamically changing uncertainty conditions.

RECOMMENDATION

Based on the empirical findings, cyber-physical system designers, engineers, operators, cybersecurity professionals, and decision-system developers should prioritize the systematic integration of uncertainty-aware mechanisms into real-time sensing, estimation, prediction, control, fault management, and recovery processes rather than relying exclusively on deterministic decision outputs. Particular emphasis should be placed on calibrated uncertainty estimation because the reduction in expected calibration error from 0.086 to 0.031 was accompanied by substantial improvements in predictive accuracy, fault detection, reliability, safety, and resilience. CPS implementations should continuously quantify uncertainty arising from sensor noise, missing observations, communication degradation, distribution shift, component deterioration, workload pressure, and cyberattack activity and should incorporate these uncertainty estimates into operational decision thresholds. The identified severity ranges should also inform adaptive control policies, with increased attention when sensor noise approaches 12%, packet loss approaches 8%, communication latency approaches 50 ms, component degradation approaches 15%, or attack intensity approaches the high-severity range, because performance deterioration became considerably stronger beyond these conditions. Reliability engineering should combine uncertainty estimation with fault detection, fault isolation, redundancy, reconfiguration, fallback control, and automated safe-state transitions to reduce unsafe actions and maintain essential functionality during abnormal operation. Resilience strategies should similarly emphasize early adaptation rather than relying primarily on post-disturbance recovery, since uncertainty-aware operation reduced maximum performance degradation, shortened adaptation and recovery times, and increased retained functionality. Edge, fog, and distributed architectures should be considered where real-time constraints are stringent because these configurations demonstrated

favorable latency and retained-functionality characteristics, although architecture selection should remain dependent on system scale, workload, device density, communication requirements, and available computational resources. Computational optimization should focus on reducing the additional processing and memory requirements associated with uncertainty estimation through efficient model architectures, adaptive uncertainty computation, hardware acceleration, selective sampling, and resource-aware inference without sacrificing calibration quality. Cybersecurity mechanisms should integrate uncertainty information with anomaly detection and response processes because combined cyber-physical disturbances and cyberattacks occurring under degraded communication conditions produced particularly substantial reliability and resilience losses. Model evaluation should extend beyond accuracy and include calibration error, tail latency, deadline misses, availability, fault-detection performance, unsafe-action frequency, retained functionality, recovery time, recovery success, and accumulated operational loss. Statistical evaluation should report confidence intervals and practical effect sizes alongside significance testing, while model robustness should be verified through repeated, grouped, and temporal cross-validation, independent holdout evaluation, sensitivity analysis, ablation testing, and severe stress conditions. Finally, uncertainty-aware CPS deployment should be treated as an integrated engineering process in which predictive intelligence, real-time computation, cybersecurity, fault tolerance, safety constraints, and resilience mechanisms are jointly optimized, thereby supporting dependable decision-making across increasingly heterogeneous and disturbance-prone cyber-physical environments.

LIMITATIONS

This study had several limitations that should be considered when interpreting the empirical findings on uncertainty-aware real-time decision-making for reliable and resilient cyber-physical systems. Although the analytical dataset contained 7,846,320 retained observations from 15 datasets, 42 CPS environments, and 3,600 experimental runs, the evaluated environments could not represent every configuration, operating condition, hardware platform, communication protocol, sensing technology, control architecture, and application encountered in real-world CPS deployments. The inclusion of smart-grid, industrial-control, autonomous, robotic, and networked critical-infrastructure environments provided substantial heterogeneity, but domain-specific operational requirements may limit direct generalization of individual numerical results to systems with substantially different safety constraints, temporal requirements, or physical dynamics. The controlled experimental, simulation, hardware-in-the-loop, and dataset-based conditions also enabled systematic manipulation of uncertainty and disturbances but could not fully reproduce the unpredictability of continuously operating physical infrastructure. Sensor noise, missing observations, packet loss, communication latency, component degradation, workload pressure, and attack intensity were organized into predefined severity ranges, whereas real operational uncertainty can change irregularly, occur simultaneously across multiple components, and exhibit temporal dependencies that are difficult to reproduce experimentally. The cybersecurity assessment included false-data injection, denial-of-service, replay, spoofing, sensor manipulation, actuator compromise, communication disruption, and combined cyber-physical events; however, these scenarios did not represent every possible attack technique, attacker behavior, attack duration, or coordinated intrusion sequence. The comparison between deterministic and uncertainty-aware approaches also necessarily simplified differences among individual algorithms because Bayesian methods, probabilistic estimators, ensemble approaches, Monte Carlo techniques, and calibrated predictive models can have substantially different computational and statistical characteristics. Computational measurements were influenced by the selected processing platforms, workload configurations, model implementations, and experimental resource allocations; therefore, the reported latency, throughput, processor utilization, memory consumption, and communication overhead should not be interpreted as universal hardware-independent benchmarks. Although cross-validation, independent holdout testing, external evaluation, bootstrap confidence intervals, sensitivity analysis, ablation experiments, and stress testing strengthened robustness assessment, external validation remained limited to the CPS environments available within the study design. Furthermore, statistical significance and moderate-to-large effect sizes established reliable experimental differences but did not guarantee equivalent operational benefits under every deployment configuration. Thresholds such as approximately 12% sensor noise,

8% packet loss, 50 ms communication latency, 15% component degradation, and high attack intensity were empirical thresholds derived from the evaluated conditions rather than universal CPS safety boundaries.

REFERENCES

[1]. Abbaspour, A., Mokhtari, S., Sargolzaei, A., & Yen, K. K. (2020). A survey on active fault-tolerant control systems. *Electronics*, *9*(9), 1513.

[2]. Abdulhamid, S. i. M., Abd Latiff, M. S., Madni, S. H. H., & Abdullahi, M. (2018). Fault tolerance aware scheduling technique for cloud computing environment using dynamic clustering algorithm. *Neural Computing and Applications*, *29*(1), 279-293.

[3]. Aburukba, R. O., AliKarrar, M., Landolsi, T., & El-Fakih, K. (2020). Scheduling Internet of Things requests to minimize latency in hybrid Fog–Cloud computing. *Future Generation Computer Systems*, *111*, 539-551.

[4]. Amin, A. A., & Hasan, K. M. (2019). A review of fault tolerant control systems: advancements and applications. *Measurement*, *143*, 58-68.

[5]. Andrade, E., Nogueira, B., Farias Júnior, I. d., & Araújo, D. (2021). Performance and Availability Trade-Offs in Fog–Cloud IoT Environments: E. Andrade et al. *Journal of Network and Systems Management*, *29*(1), 2.

[6]. Asikis, T., & Pournaras, E. (2020). Optimization of privacy-utility trade-offs under informational self-determination. *Future Generation Computer Systems*, *109*, 488-499.

[7]. Bagula, A., Ajayi, O., & Maluleke, H. (2021). Cyber physical systems dependability using cps-iot monitoring. *Sensors*, *21*(8), 2761.

[8]. Bhuiyan, E. A., Akhand, M. M. A., Das, S. K., Ali, M. F., Tasneem, Z., Islam, M. R., Saha, D., Badal, F. R., Ahamed, M. H., & Moyeen, S. (2020). A survey on fault diagnosis and fault tolerant methodologies for permanent magnet synchronous machines. *International Journal of Automation and Computing*, *17*(6), 763-787.

[9]. Cai, B., Zhang, Y., Wang, H., Liu, Y., Ji, R., Gao, C., Kong, X., & Liu, J. (2021). Resilience evaluation methodology of engineering systems with dynamic-Bayesian-network-based degradation and maintenance. *Reliability Engineering & System Safety*, *209*, 107464.

[10]. Cassottana, B., Shen, L., & Tang, L. C. (2019). Modeling the recovery process: A key dimension of resilience. *Reliability Engineering & System Safety*, *190*, 106528.

[11]. Castillo, D., Lakshminarayanan, V., & Rodriguez-Alvarez, M. J. (2021). MR images, brain lesions, and deep learning. *Applied Sciences*, *11*(4), 1675.

[12]. Chen, J., & Wan, Z. (2019). A compatible probabilistic framework for quantification of simultaneous aleatory and epistemic uncertainty of basic parameters of structures by synthesizing the change of measure and change of random variables. *Structural Safety*, *78*, 76-87.

[13]. Choi, H., Yu, H., & Lee, E. (2019). Latency-classification-based deadline-aware task offloading algorithm in mobile edge computing environments. *Applied Sciences*, *9*(21), 4696.

[14]. Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of clinical epidemiology*, *110*, 12-22.

[15]. Clough, R., Lohan, M. C., Ussher, S. J., Nimmo, M., & Worsfold, P. J. (2019). Uncertainty associated with the leaching of aerosol filters for the determination of metals in aerosol particulate matter using collision/reaction cell ICP-MS detection. *Talanta*, *199*, 425-430.

[16]. Coit, D. W., & Zio, E. (2019). The evolution of system reliability optimization. *Reliability Engineering & System Safety*, *192*, 106259.

[17]. Crombé, A., Fadli, D., Italiano, A., Saut, O., Buy, X., & Kind, M. (2020). Systematic review of sarcomas radiomics studies: Bridging the gap between concepts and clinical applications? *European Journal of Radiology*, *132*, 109283.

[18]. Darabseh, A., & Freris, N. M. (2019). A software-defined architecture for control of IoT cyberphysical systems. *Cluster Computing*, *22*(4), 1107-1122.

[19]. De Oliveira, C. S., Sanin, C., & Szczerbicki, E. (2019). Visual content representation and retrieval for cognitive cyber physical systems. *Procedia computer science*, *159*, 2249-2257.

[20]. Deng, H., Fannon, D., & Eckelman, M. J. (2018). Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata. *Energy and Buildings*, *163*, 34-43.

[21]. DeRousseau, M. A., Laftchiev, E., Kasprzyk, J. R., Rajagopalan, B., & Srubar III, W. V. (2019). A comparison of machine learning methods for predicting the compressive strength of field-placed concrete. *Construction and Building Materials*, *228*, 116661.

[22]. Enrico, Z., Mengfei, F., Zhiguo, Z., & Rui, K. (2019). Application of reliability technologies in civil aviation: Lessons learnt and perspectives. *Chinese Journal of Aeronautics*, *32*(1), 143-158.

[23]. Farooq, B. H., Deckard, B. E. R., Ziemba-Davis, B. M., Madsen, D. A., & Meneghini, M. R. M. (2020). Predictors of patient satisfaction following primary total knee arthroplasty: results from a traditional statistical model and a machine learning algorithm. *The Journal of Arthroplasty*, *35*(11), 3123-3130.

[24]. Fentaye, A. D., Baheta, A. T., Gilani, S. I., & Kyprianidis, K. G. (2019). A review on gas turbine gas-path diagnostics: State-of-the-art methods, challenges and opportunities. *Aerospace*, *6*(7), 83.

[25]. Ferreira, G. S., Veening-Griffioen, D. H., Boon, W. P., Moors, E. H., & van Meer, P. J. (2020). Levelling the translational gap for animal to human efficacy data. *Animals*, *10*(7), 1199.

[26]. Ferrer, B. R., & Lastra, J. L. M. (2017). Private local automation clouds built by CPS: Potential and challenges for distributed reasoning. *Advanced Engineering Informatics*, *32*, 113-125.
[27]. Foehr, M., Vollmar, J., Calà, A., Leitão, P., Karnouskos, S., & Colombo, A. W. (2017). Engineering of next generation cyber-physical automation system architectures. In *Multi-disciplinary engineering for cyber-physical production systems: Data models and software solutions for handling complex engineering projects* (pp. 185-206). Springer.
[28]. Fritz, R., & Zhang, P. (2018). Overview of fault-tolerant control methods for discrete event systems. *IFAC-PapersOnLine*, *51*(24), 88-95.
[29]. Gao, Z., & Liu, X. (2021). An overview on fault diagnosis, prognosis and resilient control for wind turbine systems. *Processes*, *9*(2), 300.
[30]. Gkanatsas, E., & Krikke, H. (2020). Towards a pro-silience framework: a literature review on quantitative modelling of resilient 3PL supply chain network designs. *Sustainability*, *12*(10), 4323.
[31]. Goniewicz, K., Goniewicz, M., Włoszczak-Szubzda, A., Burkle, F. M., Hertelendy, A. J., Al-Wathinani, A., Molloy, M. S., & Khorram-Manesh, A. (2021). The importance of pre-training gap analyses and the identification of competencies and skill requirements of medical personnel for mass casualty incidents and disaster training. *BMC public health*, *21*(1), 114.
[32]. Han, S., & Coulibaly, P. (2017). Bayesian flood forecasting methods: A review. *Journal of Hydrology*, *551*, 340-351.
[33]. Haque, M. A., Shetty, S., Gold, K., & Krishnappa, B. (2021). Realizing cyber-physical systems resilience frameworks and security practices. In *Security in cyber-physical systems: foundations and applications* (pp. 1-37). Springer.
[34]. Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big data*, *6*(1), 44.
[35]. Hossain, N. U. I., Jaradat, R., Hosseini, S., Marufuzzaman, M., & Buchanan, R. K. (2019). A framework for modeling and assessing system resilience using a Bayesian network: A case study of an interdependent electrical infrastructure system. *International Journal of Critical Infrastructure Protection*, *25*, 62-83.
[36]. Hua, L., Zhang, J., Li, D., & Xi, X. (2021). Fault-tolerant active disturbance rejection control of plant protection of unmanned aerial vehicles based on a spatio-temporal RBF neural network. *Applied Sciences*, *11*(9), 4084.
[37]. Huang, C., Naghdy, F., Du, H., & Huang, H. (2019). Fault tolerant steer-by-wire systems: An overview. *Annual Reviews in Control*, *47*, 98-111.
[38]. Huang, J.-C., Ko, K.-M., Shu, M.-H., & Hsu, B.-M. (2020). Application and comparison of several machine learning algorithms and their integration models in regression problems. *Neural Computing and Applications*, *32*(10), 5461-5469.
[39]. Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine learning*, *110*(3), 457-506.
[40]. Hunziker, S., & Blankenagel, M. (2021). Cross-sectional research design. In *Research design in business and management: A practical guide for students and researchers* (pp. 187-199). Springer.
[41]. Ide, T., Brzoska, M., Donges, J. F., & Schleussner, C.-F. (2020). Multi-method evidence for when and how climate-related disasters contribute to armed conflict risk. *Global Environmental Change*, *62*, 102063.
[42]. Jhawar, R., & Piuri, V. (2017). Fault tolerance and resilience in cloud computing environments. In *Computer and information security handbook* (pp. 155-173). Elsevier.
[43]. Jiang, C., Zheng, J., & Han, X. (2018). Probability-interval hybrid uncertainty analysis for structures with both aleatory and epistemic uncertainties: a review: C. Jiang et al. *Structural and Multidisciplinary Optimization*, *57*(6), 2485-2502.
[44]. Karve, P. M., Guo, Y., Kapusuzoglu, B., Mahadevan, S., & Haile, M. A. (2020). Digital twin approach for damage-tolerant mission planning under uncertainty. *Engineering Fracture Mechanics*, *225*, 106766.
[45]. Kathiravelu, P., Van Roy, P., & Veiga, L. (2019). SD-CPS: software-defined cyber-physical systems. Taming the challenges of CPS with workflows at the edge. *Cluster Computing*, *22*(3), 661-677.
[46]. Kazi Rakib Hasan, S., & Adnan, S. (2022). Quantitative Assessment of Machine Learning-Based Schedule Risk Prediction for Reducing Delays in Complex Technology Projects. *Review of Applied Science and Technology*, *1*(04), 464–501. https://doi.org/10.63125/dn8nkm85
[47]. Khaghani, F., & Jazizadeh, F. (2020). mD-resilience: A multi-dimensional approach for resilience-based performance assessment in urban transportation. *Sustainability*, *12*(12), 4879.
[48]. Kim, J. H., Hornibrook, C., & Yim, M.-S. (2020). The impact of below detection limit samples in residual risk assessments for decommissioning nuclear power plant sites. *Journal of environmental radioactivity*, *222*, 106340.
[49]. Kontopantelis, E., Sperrin, M., Mamas, M. A., & Buchan, I. E. (2018). Investigating heterogeneity of effects and associations using interaction terms. *Journal of clinical epidemiology*, *93*, 79-83.
[50]. Kunnappilly, A., Marinescu, R., & Seceleanu, C. (2018). Assuring intelligent ambient assisted living solutions by statistical model checking. International Symposium on Leveraging Applications of Formal Methods,
[51]. Kurth, M., Kozlowski, W., Ganin, A., Mersky, A., Leung, B., Dykes, J., Kitsak, M., & Linkov, I. (2020). Lack of resilience in transportation networks: Economic implications. *Transportation Research Part D: Transport and Environment*, *86*, 102419.
[52]. Lee, M., & Basu, D. (2018). An integrated approach for resilience and sustainability in geotechnical engineering. *Indian Geotechnical Journal*, *48*(2), 207-234.
[53]. Li, W., Fan, X., Huang, F., Chen, W., Hong, H., Huang, J., & Guo, Z. (2020). Uncertainties analysis of collapse susceptibility prediction based on remote sensing and GIS: Influences of different data-based models and connections between collapses and environmental factors. *Remote Sensing*, *12*(24), 4134.
[54]. Linkov, I., & Kott, A. (2019). Fundamental concepts of cyber resilience: Introduction and overview. In *Cyber resilience of systems and networks* (pp. 1-25). Springer.

[55]. Liu, Z., & Wang, J. (2020). Human-cyber-physical systems: concepts, challenges, and research opportunities. *Frontiers of Information Technology & Electronic Engineering*, *21*(11), 1535-1553.
[56]. Lork, C., Li, W.-T., Qin, Y., Zhou, Y., Yuen, C., Tushar, W., & Saha, T. K. (2020). An uncertainty-aware deep reinforcement learning framework for residential air conditioning energy management. *Applied Energy*, *276*, 115426.
[57]. M. Mabkhot, M., Ferreira, P., Maffei, A., Podržaj, P., Mądziel, M., Antonelli, D., Lanzetta, M., Barata, J., Boffa, E., & Finžgar, M. (2021). Mapping industry 4.0 enabling technologies into united nations sustainability development goals. *Sustainability*, *13*(5), 2560.
[58]. Ma, Y., Hall, D. L., Ngo, L. H., Liu, Q., Bain, P. A., & Yeh, G. Y. (2021). Efficacy of cognitive behavioral therapy for insomnia in breast cancer: a meta-analysis. *Sleep medicine reviews*, *55*, 101376.
[59]. Mahbubul, I., Mahmudul, I., Shifa, T., Mahfuza, T., & Shima Ali, S. (2022). Exploring The Potential of Visual Narratives and Interactive Gamified Communication Systems to Prevent Disease, Counter Health Misinformation, And Strengthen Public Trust in Low-Literacy Communities: A Literature Review. *American Journal of Interdisciplinary Studies*, *3*(02), 165-220. https://doi.org/10.63125/zmqs4n64
[60]. Mashfiqur, R., & Md, A. (2022). Monte Carlo Simulation and Stochastic Cash-Flow Modeling for Reducing Liquidity Risk in Project-Based Enterprises. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 382–431. https://doi.org/10.63125/zf3fwv21
[61]. McKeand, A. M., Gorguluarslan, R. M., & Choi, S.-K. (2021). Stochastic analysis and validation under aleatory and epistemic uncertainties. *Reliability Engineering & System Safety*, *205*, 107258.
[62]. Md Ashfaq, S., & Abdullah Mohammad, S. (2022). CNT/PANI Composite Electrodes for Flexible Fiber-Shaped Dye-Sensitized Solar Cells: Synthesis, Characterization, and Photovoltaic Performance. *American Journal of Interdisciplinary Studies*, *3*(02), 221-274. https://doi.org/10.63125/m334mf41
[63]. Md. Abdur, R., & Iftekhar, A. (2021). Customer Retention Forecasting in Mobile Wallet Services Using Neural Networks: A Comparative Quantitative Study. *International Journal of Business and Economics Insights*, *1*(4), 70-102. https://doi.org/10.63125/dyrpc387
[64]. Mehrdad, S., Mousavian, S., Madraki, G., & Dvorkin, Y. (2018). Cyber-physical resilience of electrical power systems against malicious attacks: A review. *Current Sustainable/Renewable Energy Reports*, *5*(1), 14-22.
[65]. Mervin, L. H., Johansson, S., Semenova, E., Giblin, K. A., & Engkvist, O. (2021). Uncertainty quantification in drug design. *Drug discovery today*, *26*(2), 474-489.
[66]. Min-Allah, N., Qureshi, M. B., Alrashed, S., & Rana, O. F. (2019). Cost efficient resource allocation for real-time tasks in embedded systems. *Sustainable Cities and Society*, *48*, 101523.
[67]. Molina, E., & Jacob, E. (2018). Software-defined networking in cyber-physical systems: A survey. *Computers & electrical engineering*, *66*, 407-419.
[68]. Nait Belaid, Y., Coudray, P., Sanchez-Torres, J., Fang, Y.-P., Zeng, Z., & Barros, A. (2021). Resilience quantification of smart distribution networks—A bird's eye view perspective. *Energies*, *14*(10), 2888.
[69]. Nan, C., & Sansavini, G. (2017). A quantitative method for assessing resilience of interdependent infrastructures. *Reliability Engineering & System Safety*, *157*, 35-53.
[70]. Nouman, H., & Alfandari, R. (2020). Identifying children suspected for maltreatment: The assessment process taken by healthcare professionals working in community healthcare services. *Children and youth services review*, *113*, 104964.
[71]. Oberkampf, W. L. (2019). Simulation accuracy, uncertainty, and predictive capability: A physical sciences perspective. In *Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives* (pp. 69-97). Springer.
[72]. Olson, L. M., Campbell, K. A., Cook, L., & Keenan, H. T. (2018). Social history: A qualitative analysis of child abuse pediatricians' consultation notes. *Child abuse & neglect*, *86*, 267-277.
[73]. Pan, Y., Jing, Y., Wu, T., & Kong, X. (2021). An integrated data and knowledge model addressing aleatory and epistemic uncertainty for oil condition monitoring. *Reliability Engineering & System Safety*, *210*, 107546.
[74]. Park, S.-T., Li, G., & Hong, J.-C. (2020). A study on smart factory-based ambient intelligence context-aware intrusion detection system using machine learning. *Journal of Ambient Intelligence and Humanized Computing*, *11*(4), 1405-1412.
[75]. Patton, R. J. (2021). Fault-tolerant control. In *Encyclopedia of systems and control* (pp. 779-785). Springer.
[76]. Peng, H., Liu, C., Zhao, D., Hu, Z., & Han, J. (2020). Security evaluation under different exchange strategies based on heterogeneous CPS model in interdependent sensor networks. *Sensors*, *20*(21), 6123.
[77]. Preuveneers, D., Tsingenopoulos, I., & Joosen, W. (2020). Resource usage and performance trade-offs for machine learning models in smart environments. *Sensors*, *20*(4), 1176.
[78]. Qureshi, M. S., Qureshi, M. B., Fayaz, M., Zakarya, M., Aslam, S., & Shah, A. (2020). Time and cost efficient cloud resource allocation for real-time data-intensive smart systems. *Energies*, *13*(21), 5706.
[79]. Rahimzad, M., Moghaddam Nia, A., Zolfonoon, H., Soltani, J., Danandeh Mehr, A., & Kwon, H.-H. (2021). Performance comparison of an LSTM-based deep learning model versus conventional machine learning algorithms for streamflow forecasting. *Water Resources Management*, *35*(12), 4167-4187.
[80]. Rajula, H. S. R., Verlato, G., Manchia, M., Antonucci, N., & Fanos, V. (2020). Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina*, *56*(9), 455.
[81]. Rekkas, A., Paulus, J. K., Raman, G., Wong, J. B., Steyerberg, E. W., Rijnbeek, P. R., Kent, D. M., & van Klaveren, D. (2020). Predictive approaches to heterogeneous treatment effects: a scoping review. *BMC Medical Research Methodology*, *20*(1), 264.

[82]. Sadia, Z., & Maishar Mostari, Z. (2022). The Impact of Interactive Dashboard Visualization on Data-Driven Decision Making in Financial Services: A Qualitative Evaluation. *International Journal of Scientific Interdisciplinary Research*, *12*(03), 01–35. https://doi.org/10.63125/3jstsv75

[83]. Scheu, M. N., Kolios, A., Fischer, T., & Brennan, F. (2017). Influence of statistical uncertainty of component reliability estimations on offshore wind farm availability. *Reliability Engineering & System Safety*, *168*, 28-39.

[84]. Segalman, D. J., & Brake, M. R. (2017). Epistemic and aleatoric uncertainty in modeling. In *The Mechanics of Jointed Structures: Recent Research and Open Challenges for Developing Predictive Models for Structural Dynamics* (pp. 593-603). Springer.

[85]. Shadi, M., Abrishami, S., Mohajerzadeh, A. H., & Zolfaghari, B. (2021). Ready-time partitioning algorithm for computation offloading of workflow applications in mobile cloud computing: M. Shadi et al. *The Journal of Supercomputing*, *77*(6), 6408-6434.

[86]. Shaiful, M., Anisur, R., & Md, A. (2022). A Systematic Literature Review on the Role of Digital Health Twins in Preventive Healthcare For Personal and Corporate Wellbeing. *American Journal of Interdisciplinary Studies*, *3*(04), 1-31. https://doi.org/10.63125/negjw373

[87]. Shaker, M. H., & Hüllermeier, E. (2020). Aleatoric and epistemic uncertainty with random forests. International Symposium on Intelligent Data Analysis,

[88]. Shandiz, S. C., Foliente, G., Rismanchi, B., Wachtel, A., & Jeffers, R. F. (2020). Resilience framework and metrics for energy master planning of communities. *Energy*, *203*, 117856.

[89]. Sharafati, A., Asadollah, S. B. H. S., & Hosseinzadeh, M. (2020). The potential of new ensemble machine learning models for effluent quality parameters prediction and related uncertainty. *Process Safety and Environmental Protection*, *140*, 68-78.

[90]. Sharma, S., & Mehra, R. (2020). Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images—a comparative insight. *Journal of digital imaging*, *33*(3), 632-654.

[91]. Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., Chen, S., Liu, D., & Li, J. (2020). Performance comparison and current challenges of using machine learning techniques in cybersecurity. *Energies*, *13*(10), 2509.

[92]. Shen, B., Chilamkurti, N., Wang, R., Zhou, X., Wang, S., & Ji, W. (2018). Deadline-aware rate allocation for IoT services in data center network. *Journal of Parallel and Distributed Computing*, *118*, 296-306.

[93]. Shirzadi, A., Shahabi, H., Chapi, K., Bui, D. T., Pham, B. T., Shahedi, K., & Ahmad, B. B. (2017). A comparative study between popular statistical and machine learning methods for simulating volume of landslides. *Catena*, *157*, 213-226.

[94]. Shuang, Q., Liu, H. J., & Porse, E. (2019). Review of the quantitative resilience methods in water distribution networks. *Water*, *11*(6), 1189.

[95]. Siddiq, F., & Scherer, R. (2019). Is there a gender gap? A meta-analysis of the gender differences in students' ICT literacy. *Educational research review*, *27*, 205-217.

[96]. Soize, C. (2017). *Uncertainty quantification* (Vol. 23). Springer.

[97]. Song, X., Liu, X., Liu, F., & Wang, C. (2021). Comparison of machine learning and logistic regression models in predicting acute kidney injury: a systematic review and meta-analysis. *International journal of medical informatics*, *151*, 104484.

[98]. Tchernykh, A., Schwiegelsohn, U., Talbi, E.-g., & Babenko, M. (2019). Towards understanding uncertainty in cloud computing with risks of confidentiality, integrity, and availability. *Journal of Computational Science*, *36*, 100581.

[99]. Tomasko, J., Stupak, M., Parizkova, D., Polachova, A., Sram, R. J., Topinka, J., & Pulkrabova, J. (2021). Short-and medium-chain chlorinated paraffins in human blood serum of Czech population. *Science of The Total Environment*, *797*, 149126.

[100]. Tran, H. T., Balchanos, M., Domerçant, J. C., & Mavris, D. N. (2017). A framework for the quantitative assessment of performance-based system resilience. *Reliability Engineering & System Safety*, *158*, 73-84.

[101]. Troya, J., Moreno, N., Bertoa, M. F., & Vallecillo, A. (2021). Uncertainty representation in software models: a survey. *Software and Systems Modeling*, *20*(4), 1183-1213.

[102]. Tweneboah-Koduah, S., Tsetse, A., & Prasad, R. (2021). Quantitative estimate of infrastructure interdependence. *Wireless Personal Communications*, *118*(1), 261-280.

[103]. Ursprung, S., Beer, L., Bruining, A., Woitek, R., Stewart, G. D., Gallagher, F. A., & Sala, E. (2020). Radiomics of computed tomography and magnetic resonance imaging in renal cell carcinoma—a systematic review and meta-analysis. *European radiology*, *30*(6), 3558-3566.

[104]. Vachtsevanos, G., Lee, B., Oh, S., & Balchanos, M. (2018). Resilient design and operation of cyber physical systems with emphasis on unmanned autonomous systems. *Journal of Intelligent & Robotic Systems*, *91*(1), 59-83.

[105]. VanBommel, S., Gellert, R., Boyd, N., & Hanania, J. (2019). Empirical simulations for further characterization of the Mars Science Laboratory Alpha Particle X-ray Spectrometer: An introduction to the ACES program. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, *441*, 79-87.

[106]. Verleysen, E., Waegeneers, N., Brassinne, F., De Vos, S., Jimenez, I. O., Mathioudaki, S., & Mast, J. (2020). Physicochemical characterization of the pristine E171 food additive by standardized and validated methods. *Nanomaterials*, *10*(3), 592.

[107]. Villar, E., Merino, J., Posadas, H., Henia, R., & Rioux, L. (2020). Mega-modeling of complex, distributed, heterogeneous cps systems. *Microprocessors and microsystems*, *78*, 103244.

[108]. Wang, H., Dey, K. K., Chen, P.-C., Li, Y., Niu, M., Cho, J.-H., Wang, X., Bai, B., Jiao, Y., & Chepyala, S. R. (2020). Integrated analysis of ultra-deep proteomes in cortex, cerebrospinal fluid and serum reveals a mitochondrial signature in Alzheimer's disease. *Molecular neurodegeneration*, *15*(1), 43.

[109]. Wang, L., Cohen, J. C., Devasenapathy, N., Hong, B. Y., Kheyson, S., Lu, D., Oparin, Y., Kennedy, S. A., Romerosa, B., & Arora, N. (2020). Prevalence and intensity of persistent post-surgical pain following breast cancer surgery: a systematic review and meta-analysis of observational studies. *British journal of anaesthesia*, *125*(3), 346-357.

[110]. Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2020). A field-wide systematic review and meta-analysis of putative risk and protective factors for radicalization outcomes. *Journal of quantitative criminology*, *36*(3), 407-447.

[111]. Wong, F., Suzuki, G., Michinaka, C., Yuan, B., Takigami, H., & de Wit, C. A. (2017). Dioxin-like activities, halogenated flame retardants, organophosphate esters and chlorinated paraffins in dust from Australia, the United Kingdom, Canada, Sweden and China. *Chemosphere*, *168*, 1248-1256.

[112]. Zaidan, A., Zaidan, B., Alsalem, M., Albahri, O. S., Albahri, A. S., & Qahtan, M. (2020). Multi-agent learning neural network and Bayesian model for real-time IoT skin detectors: a new evaluation and benchmarking methodology. *Neural Computing and Applications*, *32*(12), 8315-8366.

[113]. Zhang, H., Yuan, H., Li, G., & Lin, Y. (2018). Quantitative resilience assessment under a tri-stage framework for power systems. *Energies*, *11*(6), 1427.

[114]. Zwanenburg, A. (2019). Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *European journal of nuclear medicine and molecular imaging*, *46*(13), 2638-2655.