A Survey of Low-Probability High-Impact Events: Knowledge-Based, Data-Driven Risk Assessment in Complex Systems

Abstract

This survey paper presents a comprehensive analysis of the integration of knowledge-based and data-driven approaches in assessing low-probability high-impact events. These rare occurrences, despite their infrequency, can cause significant disruptions across various domains, necessitating advanced risk assessment methodologies. The paper highlights the importance of combining expert insights with empirical data to develop predictive models that are both accurate and reliable. It explores the application of machine learning techniques and innovative data-driven methodologies, emphasizing their role in enhancing predictive accuracy and managing uncertainties in complex systems. The survey also addresses challenges such as computational complexity, data quality, and model interpretability, proposing interdisciplinary collaborations as a pathway to overcoming these barriers. Case studies across sectors like healthcare, climate science, and power systems demonstrate the practical applications and effectiveness of these integrated approaches. The findings underscore the potential of these methodologies to improve decision-making and policy development, ultimately enhancing the resilience and adaptability of complex systems to low-probability high-impact events.

1 Introduction

1.1 Significance of Low-Probability High-Impact Events

Low-probability high-impact (HILP) events, despite their rarity, can lead to substantial disruptions across multiple domains, necessitating a thorough understanding for effective management and mitigation. For instance, the potential collapse of the Atlantic meridional overturning circulation (AMOC) exemplifies the critical implications such events can have on global climate stability [1]. In public safety, the advancement of communication technologies is vital for first responders to adequately prepare for unprecedented emergencies, underscoring the importance of readiness in the face of these rare occurrences [2].

The COVID-19 pandemic revealed significant vulnerabilities in the medical supply chain, illustrating how HILP events can disrupt essential services and necessitate robust risk management strategies [3]. In healthcare, early predictions of mortality and Length of Stay (LOS) in Intensive Care Units (ICUs) highlight the need for precise predictive models to enhance patient outcomes [4].

Complex social systems often exhibit right-skewed distributions, as seen in political conflicts and economic networks, where rare but impactful events challenge traditional risk assessment models [5]. The challenges faced by startups, characterized by the rarity of success and numerous influencing factors, further illustrate the complexities associated with HILP events [6].

In the energy sector, HILP events in power systems hold significant societal implications despite their low probability [7]. Microgrids are crucial for providing backup power during utility grid interruptions, with their resilience dependent on event type, severity, and site-specific risk profiles

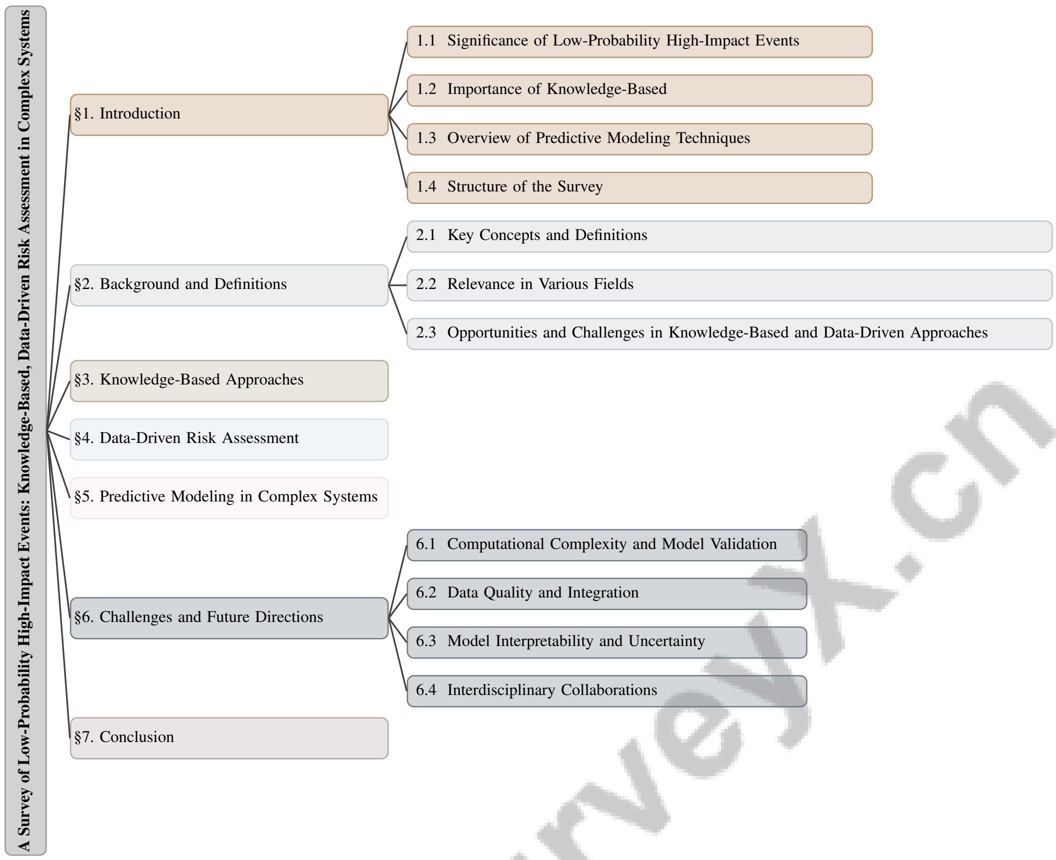


Figure 1: chapter structure

[8]. In regions like China, frequently impacted by typhoons, the electrical grid faces severe threats from destructive weather, leading to substantial economic losses [9]. Additionally, the volatility of renewable energy generation and high demand periods in electricity markets exemplify these events, exposing consumers to significant costs [10].

Extreme weather events (EWEs), such as hurricanes and atmospheric rivers, are prominent HILP events that require effective identification and tracking to mitigate their impacts [11]. The catastrophic Black Summer bushfires in Australia (2019-2020) further illustrate the extensive damage these events can inflict on communities and ecosystems [12]. Understanding and preparing for such events is paramount, as they complicate decision-making and risk assessments across various sectors, highlighting the need for strategic resource allocation for disaster preparedness [13].

1.2 Importance of Knowledge-Based, Data-Driven Approaches

The integration of knowledge-based and data-driven approaches is vital for effective risk assessment, especially concerning HILP events. This hybrid methodology combines expert insights with empirical data, facilitating the development of predictive models that are both accurate and reliable. The urgency for breakthroughs in predicting tipping points and simulating non-stationary dynamics in complex systems emphasizes the importance of this integration [14]. The proposed explainable case-based reasoning (CBR) approach underscores the necessity for transparent decision-making in financial services, highlighting the role of merging knowledge-based and data-driven methods [15].

In healthcare, the fusion of expert clinical judgment with multivariate temporal data from Electronic Health Records (EHRs) enhances risk assessment accuracy, as demonstrated by LSTM-based methods predicting ICU mortality [4]. A hybrid approach combining data-driven and physics-based modeling is crucial for overcoming limitations in accurately representing dynamic behaviors [16]. Furthermore, the need for such an approach is evident in providing theoretical guarantees for stability and robustness in controlling unknown systems [17].

The application of data-driven stochastic predictive control schemes, which formulate an Optimal Control Problem (OCP) using historical state, input, and disturbance trajectories, illustrates the significance of this integration for forecasting future behaviors [18]. The underestimation of low-probability high-impact events in resource allocation highlights the necessity for a comprehensive approach that merges knowledge-based insights with data-driven analysis [13].

The combination of knowledge-based and data-driven approaches establishes a robust framework for risk assessment, enabling more accurate predictions and effective management of HILP events. This integrated methodology significantly enhances model predictive accuracy while addressing the complexities inherent in real-world systems through applications in data journalism, model-driven analytics, and prescriptive machine learning. Advanced statistical techniques and frameworks, such as Latent Dirichlet Allocation for topic modeling and the cross-industry standard process for data mining, facilitate the extraction of meaningful insights from extensive datasets, ensuring robustness and reliability in analytics across various fields, including healthcare, finance, and network security [19, 20, 21, 22, 23].

1.3 Overview of Predictive Modeling Techniques

Predictive modeling techniques are essential for risk assessment, particularly in the context of HILP events. These techniques utilize historical data and advanced algorithms to forecast future outcomes, enabling proactive risk management strategies. Machine learning approaches, such as Data-driven modeling for battery cycle life prediction (DDM-BCLP), exemplify the application of early cycle discharge voltage curves to predict battery life, demonstrating the potential of these models in energy systems risk assessment [24].

In environmental hazards, the Bushfire Severity Predictive Model (BSPM) integrates Landsat imagery with spectral indices and topographical and climatic factors to accurately predict bushfire severity, showcasing predictive models' role in managing natural disasters [12]. Similarly, Long Short-Term Memory (LSTM) networks have been effectively employed in healthcare to analyze early physiological data, facilitating predictions of patient outcomes such as mortality and LOS in ICUs [4].

These examples illustrate the extensive range and flexibility of predictive modeling techniques across diverse fields such as data journalism, healthcare, and manufacturing. For instance, advanced statistical methods in data journalism analyze complex datasets like the WikiLeaks Afghanistan war logs, revealing insights into fatality rates related to specific conflict circumstances. Model-driven analytics integrates domain knowledge with data from sectors such as IoT and healthcare, addressing the challenges posed by large volumes of intricate data. In medicine, machine learning enhances traditional predictive methods, enabling the development of real-time clinical decision-making tools for various healthcare scenarios. These instances highlight how predictive modeling adapts to meet specific challenges across domains, driving innovation and insight [19, 25, 23, 22]. By incorporating data-driven and knowledge-based insights, these models improve the accuracy and reliability of risk assessments, ultimately contributing to informed decision-making in complex systems.

1.4 Structure of the Survey

This survey is organized into seven primary sections, each addressing critical aspects of low-probability high-impact events and the methodologies employed for their assessment and management. Section 1 introduces the significance of these events, emphasizing the necessity of knowledge-based, data-driven approaches for effective risk assessment in complex systems, and provides an overview of predictive modeling techniques utilized in this context. Section 2 delves into the background and definitions, establishing a foundational understanding of key concepts such as knowledge-based approaches, data-driven risk assessment, complex systems, and predictive modeling, while exploring their relevance across various fields and highlighting both opportunities and challenges.

Section 3 focuses on knowledge-based approaches, examining the role of expert insights in managing risks associated with these rare events. It discusses methodologies for integrating expert knowledge into predictive models and risk assessments, supported by case studies demonstrating successful applications. Section 4 shifts attention to data-driven risk assessment, exploring the application of data analytics and machine learning techniques in identifying patterns and predicting rare events

AI-generated, for reference only.

within complex systems. This section also reviews existing frameworks and platforms that facilitate data-driven risk assessment.

In Section 5, the development and application of predictive models for anticipating HILP events are discussed. This section covers various modeling techniques, including statistical models, machine learning algorithms, and hybrid approaches, while emphasizing the importance of model validation and uncertainty quantification. Section 6 identifies key challenges in implementing these methodologies, proposing future research directions to enhance model accuracy, data integration, and interpretability, while suggesting potential interdisciplinary collaborations to address these challenges.

In conclusion, Section 7 synthesizes the key insights from the survey, emphasizing the critical importance of integrating knowledge-based and data-driven methodologies for effective risk assessment. It highlights the necessity for a systematic approach, such as the CRISP-DM framework, to address challenges associated with data-driven knowledge discovery models in various industries. By showcasing advanced statistical methods, as demonstrated in the analysis of the WikiLeaks Afghanistan war logs, the section illustrates how these integrated approaches can enhance understanding and decision-making in complex environments, ultimately leading to more robust and reliable risk assessments [19, 23]. The conclusion underscores the potential impact of these methodologies on decision-making and policy development within complex systems.The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Key Concepts and Definitions

Low-probability high-impact (HILP) events, despite their rarity, can disrupt various sectors significantly, necessitating a robust framework for understanding and management [7]. This framework hinges on key concepts such as knowledge-based approaches, data-driven risk assessment, complex systems, and predictive modeling. Knowledge-based approaches utilize expert insights and established methodologies crucial for decision-making in areas like climate modeling and energy systems, where expert interpretation of complex phenomena is essential [26]. These approaches are indispensable for predictive maintenance and survival modeling, ensuring system reliability through thorough risk assessment [16].

Data-driven risk assessment, employing advanced analytics and machine learning, identifies patterns and predicts rare events, excelling at processing large datasets to uncover insights missed by traditional methods [15]. The shift from physics-based to data-driven techniques is evident in fields like uncontrolled object re-entry prediction, where predictive accuracy is significantly enhanced [27]. The Data-Driven Vendor Management (DDVM) method exemplifies this adaptability, using historical data to forecast customer behavior [18].

Complex systems, characterized by intricate interdependencies and dynamic interactions, pose challenges for modeling and simulation. Uncovering hidden structures and dynamics is vital for applications in quantitative biology and climate simulation [14]. The under-representation of HILP events in climate projections, such as sea-level rise, highlights the limitations of ensemble models and the need for more comprehensive approaches [13].

Predictive modeling techniques, including machine learning and hybrid models, anticipate complex system behaviors under various scenarios, providing insights into potential outcomes and supporting proactive risk management. This is exemplified in predicting battery cycle life and the transformative impact of AI across sectors [28]. Emphasizing predictive accuracy is crucial in complex systems, where data-driven model augmentation frameworks enhance reliability [4].

In controlling unknown systems, concepts such as data-driven model predictive control (MPC), stability, robustness, and constraint satisfaction are vital for ensuring accurate predictions while maintaining system stability and adhering to operational constraints [17]. Integrating knowledge-based and data-driven methodologies creates a robust framework for risk assessment, enabling precise predictions and effective management of HILP events, capturing the complexities of real-world systems, and facilitating informed decision-making and strategic planning [16].

2.2 Relevance in Various Fields

HILP events have profound implications across sectors, necessitating nuanced understanding for effective mitigation. In public health, emerging infectious diseases pose significant threats in low-income countries, contributing to high mortality rates [29]. This underscores the need for robust data-driven frameworks to enhance disease surveillance and control. In environmental management, urban planning, and disaster response, understanding bushfire severity is critical, especially in regions like Australia, where such events can devastate ecosystems and communities [12]. Predictive modeling in these areas facilitates proactive strategies, enabling stakeholders to anticipate and respond to threats effectively.

The energy sector faces challenges from HILP events, such as widespread grid failures and extreme weather, which can disrupt power supply and have severe economic impacts. These events are often inadequately represented in traditional risk assessments due to their rarity, necessitating specialized analytical frameworks for evaluating potential impacts and informing resource allocation strategies [7, 13]. Enhancing the resilience of power systems, including microgrids, is crucial for maintaining energy stability and ensuring rapid recovery from disruptions. Data-driven methodologies can bolster infrastructure robustness against these rare events.

In finance, HILP events like market crashes and systemic failures complicate risk management frameworks due to their unpredictable nature and societal repercussions. These events require specialized analytical approaches that account for complex interplays of diverse threats and uncertainties, similar to frameworks in power systems and supply networks. Understanding HILP events is essential for enhancing financial system resilience and mitigating risks [30, 7]. Knowledge-based and data-driven methods aid in identifying early warning signals and developing strategies to safeguard financial stability.

Comprehending HILP events is vital across fields, as events like major power outages or catastrophic natural disasters can lead to significant societal consequences despite their infrequency. Research indicates that human responses to HILP events are influenced by personal histories and heuristics, such as the availability heuristic, which can skew risk perceptions based on recent experiences. Developing qualitative frameworks for analyzing HILP events in complex systems, like power infrastructure, highlights the need for tailored analytical approaches to address unique uncertainties and decision-making challenges. Interdisciplinary efforts to enhance communication and understanding of HILP events are crucial for effective risk management and intervention strategies [31, 7]. Comprehensive risk assessment methodologies that combine knowledge-based insights with data-driven techniques enable stakeholders to better prepare for and manage the challenges posed by HILP events, thereby safeguarding societal well-being and promoting sustainable development.

2.3 Opportunities and Challenges in Knowledge-Based and Data-Driven Approaches

Integrating knowledge-based and data-driven methodologies in risk assessment offers significant opportunities for enhancing predictive accuracy and strategic decision-making across sectors. These hybrid approaches establish a framework for improving resilience and adaptability of complex systems to HILP events. To maximize the benefits of innovative methodologies and data-driven approaches in evaluating startups, analyzing HILP events in power systems, and enhancing digital asset management with Graph Neural Networks, challenges such as accurate data interpretation, balancing computational efficiency with analytical precision, and implementing robust frameworks to accommodate inherent complexities and uncertainties must be addressed [32, 6, 19, 7].

A primary challenge in data-driven approaches is the high dimensionality and noise in sensor data, along with time-dependence in system degradation, complicating the modeling process [28]. Additionally, data quality, particularly in Electronic Health Records (EHRs), poses hurdles due to missing or imbalanced data, adversely affecting predictive models like Long Short-Term Memory (LSTM) networks [4]. Risks of naïve extrapolation, data sampling biases, and interpreting statistical associations as causal relationships further complicate data-driven methodology implementation [33].

From a methodological standpoint, deriving rigorous bounds on estimation errors and ensuring closed-loop stability are critical challenges that must be addressed to enhance data-driven model reliability [17]. The complexities introduced by noise necessitate robust frameworks capable of accommodating uncertainties.

Despite these challenges, innovative methodologies offer promising solutions. For instance, integrating data-driven models into existing physics-based frameworks can yield more accurate predictions by addressing uncertainties in real-world processes [16]. Improved fault detection in dynamical systems is exemplified by the Occupation Kernel Principal Component Analysis (OKPCA) method, which enhances robustness against noise and accommodates irregularly sampled data [34].

3 Knowledge-Based Approaches

Risk assessment for low-probability high-impact (HILP) events greatly benefits from knowledge-based approaches, which incorporate behavioral science insights to refine predictive models and decision-making processes. This section delves into how behavioral science and cognitive biases enhance risk assessments.

3.1 Integration of Behavioral Science and Cognitive Biases

Incorporating behavioral science and cognitive biases into risk assessments is crucial for effectively managing HILP events. Cognitive biases such as overconfidence, availability heuristics, and anchoring can skew risk perception, leading to suboptimal decision-making. Integrating advanced statistical techniques with insights from data-driven journalism and computational social science significantly enhances predictive accuracy and reliability. For example, the analysis of WikiLeaks Afghanistan war logs demonstrates the utility of these models in addressing uncertainties in climate projections and the limitations of small sample data [19, 35, 36].

Frameworks like Identification for Control (I4C) optimize predictive models using control performance data, ensuring outputs align with expert expectations [37]. Techniques such as Latent Dirichlet Allocation (LDA) facilitate the integration of expert insights into risk assessments [19]. Data-driven closure models exemplify this integration by combining parametric formulations with stochastic discrepancy tensors, thus enhancing model robustness and accommodating uncertainties [38]. A bilevel framework for decision-making under uncertainty further emphasizes aligning AI-generated outputs with human understanding for informed decision-making [39].

Theoretical perspectives from decision theory advocate for resource allocation based on expected utility, integrating behavioral science into risk assessments [13]. The Fundamental Lemma, applied in data-driven control strategies, provides theoretical guarantees for predictive and stable models [17]. Aligning AI-driven discovery processes with human cognitive frameworks is essential for effectively incorporating behavioral science principles into risk assessment methodologies [40].

By addressing how personal experiences and heuristics influence decision-making, integrating behavioral science and cognitive biases refines predictive models for HILP events. This approach underscores the importance of understanding human behavior nuances, such as the tendency to rely on recent experiences over statistical warnings. Consequently, it enhances prediction accuracy in fields like climate risk management and data journalism [31, 35, 19, 41]. Acknowledging human factors in risk perception and decision-making leads to more accurate and actionable insights, improving the management of complex systems.

3.2 Case Studies and Applications

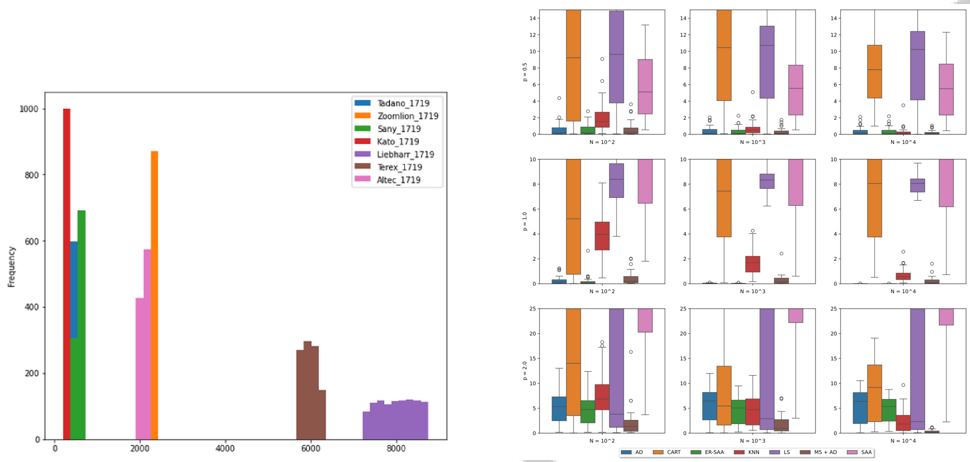
Case studies across various domains illustrate the effectiveness of knowledge-based approaches in managing HILP events. In climate science, hybrid models like ResMLP effectively simulate complex systems, enhancing predictive accuracy and managing climate-related risks [42].

In power systems, Graph Signal Processing (GSP) techniques improve tasks such as denoising and anomaly detection, strengthening infrastructure resilience against high-impact events. Simulations using real data from typhoon scenarios, such as typhoon 'Mangkhut', demonstrate the efficacy of these methods [43, 9].

The healthcare sector benefits from knowledge-based approaches, as shown by a study using the MIMIC-III dataset with Long Short-Term Memory (LSTM) networks to predict in-hospital mortality and Length of Stay (LOS), integrating clinical knowledge with data-driven methods for improved patient outcomes [4]. In infectious disease control, a two-stage optimization framework utilizes data-driven techniques to model epidemic progression and optimize resource allocation, applying knowledge-based strategies to enhance public health responses [29].

In climate data analysis, the Variational Autoencoder Transformer (VAEformer) method showcases the use of dual Variational Autoencoder architectures for compressing large datasets, retaining essential information for scientific analysis [44].

These case studies highlight the adaptability and effectiveness of knowledge-based methodologies in improving predictive accuracy and system resilience against HILP events. For instance, a qualitative framework for analyzing HILP events in power systems emphasizes understanding diverse threats and uncertainties, while a genetic model revision framework for river water quality modeling shows how integrating prior knowledge with data-driven approaches enhances model accuracy and efficiency [7, 45].



(a) Frequency Distribution of Different Vehicles[46]

(b) Comparison of Box Plots for Different Algorithms and Data Sizes[47]

Figure 2: Examples of Case Studies and Applications

As depicted in Figure 2, case studies and applications demonstrate the practical utility of theoretical models in knowledge-based approaches. The analysis of vehicle frequency distribution and algorithm performance through visual data representations illustrates this. The "Frequency Distribution of Different Vehicles" case study uses a bar chart to show trends across vehicle categories, aiding industries reliant on vehicle availability. The "Comparison of Box Plots for Different Algorithms and Data Sizes" provides a comparative analysis tool, showing metric distribution across algorithms and data sizes, helping stakeholders evaluate algorithm effectiveness. These case studies exemplify the application of knowledge-based approaches in real-world scenarios, offering actionable insights and enhancing decision-making [46, 47].

4 Data-Driven Risk Assessment

Category	Feature	Method
Machine Learning Techniques in Risk Assessment	Initial Health Prediction Complex Relationship Modeling	LSTM[4] DDSPC[18]
Innovative Data-Driven Methodologies	Robust Analysis Predictive Modeling Efficient Computation	ER-SAA[48], DMD[49] ANTL-CDM[50], ESN[51] HN[27], NG-RC[14]
Frameworks and Platforms for Risk Assessment	Data Integration	SAELP[5], BSPM[12]

Table 1: The table presents a comprehensive overview of various machine learning techniques, innovative data-driven methodologies, and frameworks/platforms utilized in risk assessment. It categorizes these methods based on their specific features and the techniques employed, highlighting their applications in predicting and managing risks across diverse sectors. This systematic classification underscores the integration of advanced algorithms and data-driven approaches in enhancing predictive accuracy and robustness in complex systems.

Integrating advanced methodologies in data-driven risk assessment is crucial for improving predictive accuracy and managing uncertainties in complex systems. Table 1 provides a detailed categorization of machine learning techniques, innovative methodologies, and frameworks/platforms that are pivotal in advancing risk assessment practices. Additionally, Table 4 offers a comparative overview of three distinct machine learning methodologies, illustrating their predictive capabilities and application areas within the context of data-driven risk assessment. The following subsection explores the application of machine learning techniques, which are increasingly vital for analyzing extensive datasets and identifying patterns that inform risk management strategies. We will examine how these techniques enhance risk assessment across various sectors, highlighting their potential to address low-probability high-impact events.

4.1 Machine Learning Techniques in Risk Assessment

Method Name	Predictive Accuracy	System Dynamics	Uncertainty Management
NG-RC[14]	Extrapolate Tipping Points	Non-stationary Dynamics	Handle Non-stationary
LSTM[4]	Improving Prediction Accuracy	Predict Patient Outcomes	Handling Missing Values
DDSPC[18]	Improve Precision	Complex Systems	Manage Uncertainties

Table 2: Comparison of machine learning methods in risk assessment, highlighting their predictive accuracy, adaptability to system dynamics, and uncertainty management capabilities. The table presents three methods: NG-RC, LSTM, and DDSPC, each demonstrating distinct strengths in handling complex, non-stationary, and uncertain environments.

Machine learning techniques are essential in risk assessment, especially for managing low-probability high-impact events. These methods exploit data-driven insights to predict rare occurrences, thereby enhancing risk management across sectors. The NG-RC method, which uses polynomial expansion of input data and a bifurcation parameter, illustrates the potential of such techniques in predicting nonlinear system dynamics [14]. This approach is particularly useful for understanding complex system behaviors under varying conditions.

In healthcare, Long Short-Term Memory (LSTM) networks assess ICU patient outcome risks based on early physiological measurements, showcasing machine learning's capability to improve predictive accuracy in critical care [4]. The DL-NSGPR model, combining deep learning with nonstationary Gaussian process regression, enhances Remaining Useful Life (RUL) prediction accuracy, demonstrating advanced machine learning's role in predictive maintenance [28].

The Decentralized Data-Enabled Predictive Control (DeePC) framework uses input-output measurements to predict future behaviors and optimize control actions without a parametric model, highlighting machine learning's utility in providing robust control strategies under uncertainty [17]. This framework underscores machine learning's adaptability in controlling complex systems, ensuring stability and robustness.

In uncertainty quantification and propagation, the data-driven Optimal Control Problem (OCP) method incorporates polynomial chaos expansions, illustrating machine learning's role in managing uncertainties in predictive models [18]. This is crucial when accurate risk predictions depend on effectively handling uncertainties.

New techniques for estimating probabilities, accounting for discrepancies between observed frequencies and actual probabilities, enhance machine learning model reliability in risk assessment [13]. This advancement ensures predictions are accurate and reflective of real-world conditions.

Table 2 provides a comparative analysis of various machine learning techniques employed in risk assessment, emphasizing their predictive accuracy and capabilities in managing system dynamics and uncertainties. Integrating machine learning techniques into risk assessment frameworks significantly improves predictive accuracy and enables proactive risk management strategies. Transitioning from traditional predictive modeling to prescriptive modeling, these frameworks facilitate informed decision-making across contexts like healthcare and industrial production. Machine learning algorithms analyze vast datasets from electronic health records and other sources to generate precise clinical outcome predictions while addressing ethical decision-making considerations. In industries like heavy machinery, machine learning optimizes production processes and assesses risk factors, enhancing operational efficiency and competitiveness [46, 52, 21, 25]. By leveraging diverse data sources and advanced algorithms, these methodologies provide essential insights that bolster complex systems' resilience and adaptability facing low-probability high-impact events.

4.2 Innovative Data-Driven Methodologies

Method Name	Modeling Techniques	Predictive Accuracy	Scalability and Efficiency
ANTL-CDM[50]	Neural Networks	High Predictive Accuracy	Scalable Method
HN[27]	Self-attention Mechanisms	Improved Predictive Accuracy	Reduced Computational Costs
DMD[49]	Model-X Knockoffs	Lightgbm, Xgboost	-
ESN[51]	Neural Networks	Predicting Chaotic Systems	Complex Systems
NG-RC[14]	Reservoir Computing Methods	Enhanced Prediction Capabilities	Efficient Modeling
ER-SAA[48]	Prediction Model	Improve Decision-making	Limited Data Regimes

Table 3: Overview of innovative data-driven methodologies highlighting their modeling techniques, predictive accuracy, and scalability. The table summarizes various methods such as ANTL-CDM, HigeNet, DMD, ESN, NG-RC, and ER-SAA, emphasizing their contributions to enhancing predictive capabilities and computational efficiency in risk assessment frameworks. These methodologies demonstrate significant advancements in handling complex systems and improving decision-making processes.

Innovative data-driven methodologies have greatly expanded risk assessment frameworks by introducing advanced techniques that enhance predictive accuracy and robustness in complex systems. Neural networks for forecasting nighttime light (NTL) sequences exemplify this, enabling continuous urban change monitoring based on deviations from expected forecasts [50]. This method demonstrates neural networks' potential in capturing temporal dynamics and facilitating real-time urban monitoring.

Tree-based models like LightGBM and XGBoost have shown superior performance in predictive modeling, with LightGBM achieving the lowest Root Mean Square Error (RMSE) and XGBoost excelling in Mean Absolute Percentage Error (MAPE) [52]. These models exemplify tree-based algorithms' effectiveness in handling complex datasets and providing accurate predictions, making them invaluable in risk assessment.

HigeNet introduces a unique self-attention mechanism, NeuralSparse, optimizing computational resources while maintaining high predictive accuracy [27]. This innovation highlights efficient computational strategies' importance in enhancing data-driven models' scalability and applicability in risk assessment contexts.

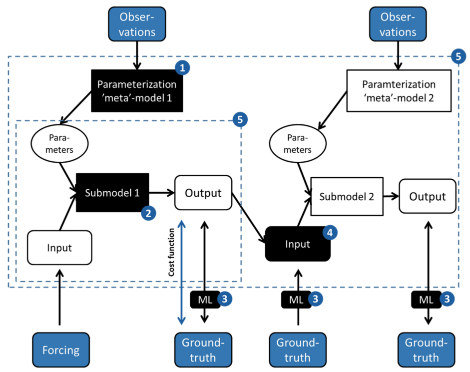
The Diamond methodology offers robust False Discovery Rate (FDR) control and versatility across machine learning models, enabling effective non-additive interaction discovery [49]. This approach is beneficial in uncovering complex data relationships, improving predictive models' interpretability and reliability.

Echo State Networks (ESNs) are powerful tools for predicting unseen dynamics, acting as early-warning systems for critical transitions [51]. ESNs' ability to generalize across unseen scenarios underscores their utility in anticipating low-probability high-impact events and enhancing proactive risk management strategies.

The NG-RC method transforms input data into a higher-dimensional space using polynomial combinations, facilitating efficient modeling of complex dynamical behaviors [14]. This technique exemplifies advanced mathematical frameworks' integration into data-driven methodologies, improving risk assessment models' robustness and adaptability.

These innovative methodologies represent significant advancements in data-driven risk assessment, providing robust tools for predicting and managing low-probability high-impact events across domains. By employing advanced methodologies and computational techniques, these approaches significantly improve risk assessment frameworks' effectiveness, enabling them to navigate real-world systems' intricacies and uncertainties, such as High-Impact Low-Probability (HILP) events in power systems and large-scale terrorist incidents. This enhancement systematically addresses diverse threats, failures, and unique uncertainties associated with these events, guiding analysts in developing tailored quantitative analyses that are robust and relevant to specific decision-making contexts [5, 19, 7, 23].

Table 3 provides a comprehensive overview of innovative data-driven methodologies employed in enhancing risk assessment frameworks, detailing their modeling techniques, predictive accuracy, and scalability.



(a) UCB on

Figure 3: Examples of Innovative Data-Driven Methodologies

As shown in Figure 3, innovative methodologies in data-driven risk assessment have emerged as powerful tools for enhancing decision-making processes. Two notable examples are depicted in Figure 3, highlighting distinct approaches to leveraging data for improved outcomes. The first example, "UCB on

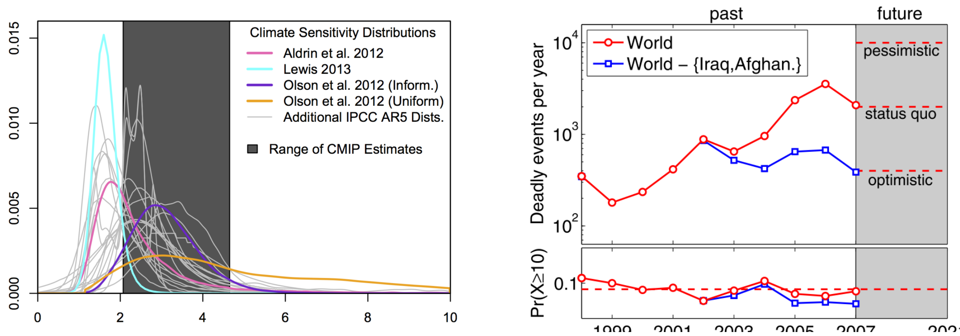
4.3 Frameworks and Platforms for Risk Assessment

Frameworks and platforms for data-driven risk assessment are vital for enhancing the prediction and management of low-probability high-impact events across domains. These systems integrate diverse data sources and advanced analytical techniques to provide comprehensive risk evaluations and informed decision-making processes. The Bushfire Severity Predictive Model (BSPM) combines remote sensing data, machine learning algorithms, and environmental factors to forecast bushfire severity across Australia [12]. This integration exemplifies data-driven frameworks' capacity to address complex environmental challenges effectively.

The Cross-Industry Standard Process for Data Mining (CRISP-DM) provides a structured approach to data-driven risk assessment, organizing methods into phases of business understanding, data understanding, data preparation, modeling, evaluation, and deployment [23]. This framework emphasizes these phases' interdependencies, ensuring each step aligns with overarching business objectives and enhances the risk assessment process's robustness and reliability. By following the CRISP-DM methodology, organizations can systematically develop and implement data-driven models that are effective and adaptable to changing conditions.

These frameworks and platforms highlight a holistic approach to risk assessment, leveraging data-driven insights and domain-specific knowledge. By combining advanced analytics with practical applications, these systems enhance the ability to identify and mitigate potential risks associated with High-Impact Low-Probability (HILP) events, such as blackouts in power systems. This integration improves complex systems' resilience and adaptability and addresses the multifaceted uncertainties and diverse threats inherent in HILP events. Developing a qualitative framework for analyzing these events and implementing user-centric anomaly detection models further supports robust decision-making and continuous performance optimization, ultimately leading to more effective management of low-probability, high-impact scenarios [20, 7].

As shown in Figure 4, data-driven approaches in risk assessment offer nuanced insights and predictions that inform strategic decision-making. The example provided highlights two distinct yet complementary frameworks and platforms for risk assessment, visualized through detailed graphs. The first graph, "Climate Sensitivity Distributions," presents a spectrum of climate sensitivity estimates from the Community Climate Model Intercomparison Project (CMIP), capturing climate models' variability and uncertainty. This visualization is crucial for understanding potential climate



(a) Climate Sensitivity Distributions[35]

(b) Past and Future Mortality Trends in the World and Iraq/Afghanistan[5]

Figure 4: Examples of Frameworks and Platforms for Risk Assessment

impacts. The second graph, "Past and Future Mortality Trends in the World and Iraq/Afghanistan," offers a historical and predictive analysis of deadly events over time, emphasizing differential trends globally and in conflict zones like Iraq and Afghanistan. Together, these examples underscore the importance of leveraging comprehensive data sets and sophisticated modeling techniques to assess and anticipate risks across diverse domains, from environmental changes to geopolitical instabilities [35, 5].

Feature	NG-RC	LSTM	DL-NSGPR
Predictive Accuracy	High	Improved	Enhanced
Methodology Type	Polynomial Expansion	Recurrent Network	Deep Learning
Application Domain	Nonlinear Dynamics	Healthcare	Predictive Maintenance

Table 4: This table presents a comparative analysis of three advanced machine learning methodologies—NG-RC, LSTM, and DL-NSGPR—highlighting their predictive accuracy, methodological frameworks, and application domains. The NG-RC method utilizes polynomial expansion for nonlinear dynamics, LSTM networks are employed in healthcare for outcome risk prediction, and DL-NSGPR enhances predictive maintenance through deep learning and Gaussian process regression. This comparison underscores the diverse applications and methodological innovations driving improvements in predictive accuracy across various sectors.

5 Predictive Modeling in Complex Systems

5.1 Statistical and Probabilistic Models

Statistical and probabilistic models are integral to predictive modeling, offering tools to navigate the uncertainties and variabilities in complex systems. These models employ statistical inference to derive insights from data, facilitating the prediction of future events based on existing patterns. In hybrid Digital Twin models, they simulate process dynamics, effectively capturing the complexities of physical systems [16]. The NG-RC method demonstrates the integration of probabilistic techniques, particularly under conditions of limited stationary data, by incorporating time-varying parameters to enhance predictive accuracy in complex environments, which is vital for managing low-probability, high-impact events [14].

In predictive maintenance, statistical models are crucial for fault detection and diagnosis. For instance, the Occupation Kernel Principal Component Analysis (OKPCA) method effectively identifies faults in nonlinear systems, such as quadrotor models with simulated actuator faults, thereby enhancing predictive maintenance reliability [34]. Data-driven Model Predictive Control (MPC) frameworks utilize statistical models to achieve performance comparable to traditional model-based approaches, particularly with noise-free data, underscoring the potential of statistical methods in optimizing control strategies [17].

In healthcare, statistical and probabilistic models provide critical insights from atypical observations, as exemplified by the CPIR method, which supports clinicians in making informed decisions based on patient data [26]. These models also play a significant role in data journalism and computational social science, where techniques like Latent Dirichlet Allocation and recursive partitioning of negative binomial distributions have been used to analyze complex datasets, such as the WikiLeaks Afghanistan war logs, revealing patterns in fatality rates. Furthermore, combining conditional logistic regression with subjective Bayesian methods enhances predictive accuracy in competitions like the Academy Awards by integrating expert opinions with historical data [19, 41]. Thus, statistical and probabilistic models significantly enhance prediction accuracy and reliability across diverse domains.

5.2 Hybrid and Semiparametric Models

Hybrid and semiparametric models are crucial in predictive modeling, merging parametric and nonparametric approaches to address inaccuracies and improve forecasting capabilities. The Semiparametric Forecasting and Filtering (SPFF) method exemplifies this integration by combining parametric models with nonparametric techniques to correct model errors, thereby enhancing prediction accuracy in dynamic environments [53]. Reservoir computing, another hybrid modeling technique, effectively approximates traditional multiscale asymptotic methods, even when scale separation is minimal, which is essential for capturing chaotic system dynamics [54].

The adaptability of hybrid models is evident in their application to climate prediction, where they adeptly capture nonlinear relationships within climate systems. By integrating data-driven approaches with prior knowledge, hybrid models enhance predictive reliability and interpretability, addressing challenges such as model errors and data limitations. This method allows for independent training of local models while promoting collaboration, leading to more accurate predictions in various applications, including risk analysis and ecological modeling [45, 36, 53, 55, 23].

5.3 Model Validation and Uncertainty Quantification

Benchmark	Size	Domain	Task Format	Metric
DL-NSGPR[28]	200	Prognostics And Health Management	Remaining Useful Life Prediction	RMSE
NeuroClim[42]	1,000,000	Climate Modeling	Hybrid Modeling	RMSE, pseudo-Radiative forcing
GDPS-SN[56]	60,000	Weather Forecasting	Forecast Verification	RMSE, ACC
UQ-PW[57]	39,000	Weather Forecasting	Probabilistic Forecasting	CRPS, CRPSS
Lachesis[20]	5,000	Networking	Time Series Forecasting	MSE, RMSE
WQP[52]	26,085	Water Quality	Regression	RMSE, MAPE

Table 5: This table presents a summary of various benchmarks used for evaluating predictive models across different domains. It details the size, domain, task format, and evaluation metrics employed, highlighting the diversity in model validation and uncertainty quantification approaches. Such benchmarks are crucial for understanding the effectiveness and reliability of predictive models in complex systems.

Model validation and uncertainty quantification are critical components of predictive modeling, especially for low-probability, high-impact events, ensuring the reliability of model outputs. These processes provide insights into the accuracy and robustness of predictive models. The CBR_E method highlights the importance of performance evaluation metrics such as accuracy, precision, recall, and F1-score for assessing model reliability [15]. The DL-NSGPR model exemplifies the integration of uncertainty quantification, delivering accurate predictions alongside uncertainty estimates, emphasizing the need to quantify uncertainties to bolster model reliability [28].

In model validation, the Diamond methodology presents a robust framework for identifying non-additive interactions from fitted machine learning models while controlling the False Discovery Rate (FDR) through a calibration process [49]. This underscores the necessity of rigorous validation techniques to maintain model integrity. Moreover, integrating physical knowledge into deep learning models enhances interpretability and plausibility, as demonstrated in Earth system modeling [33]. This integration is crucial for aligning models with known physical principles, thereby increasing their credibility.

New probability estimation methods reveal that actual probabilities may significantly exceed observed frequencies, further illustrating uncertainty quantification's role in refining predictive models [13].

This ensures models account for potential discrepancies between observed data and real-world phenomena. Incorporating model validation and uncertainty quantification into predictive frameworks enhances the robustness and reliability of predictions, addressing surrogate models' limitations in uncertainty quantification. This is particularly vital in complex domains like weather forecasting and risk analysis, where estimation errors can have significant consequences. Employing methodologies such as conformal prediction and Bayesian approaches provides guaranteed coverage for predictions, thereby improving decision-making across various applications, from industrial manufacturing to environmental forecasting [57, 58, 23, 36]. Advanced statistical methods and evaluation metrics thus establish a comprehensive foundation for managing the complexities and uncertainties inherent in complex systems, ultimately supporting informed decision-making. Table 5 provides an overview of representative benchmarks utilized in model validation and uncertainty quantification, illustrating the application of diverse evaluation metrics across various domains.

6 Challenges and Future Directions

Addressing low-probability high-impact events requires overcoming challenges such as computational complexity, model validation, data quality, and integration. These factors are crucial for developing reliable predictive models. Figure 5 illustrates the challenges and future directions in developing predictive models for these events, specifically highlighting issues related to computational complexity, data quality, model interpretability, and the necessity for interdisciplinary collaborations. The following subsections explore these challenges and their implications for enhancing model reliability and applicability in real-world scenarios.

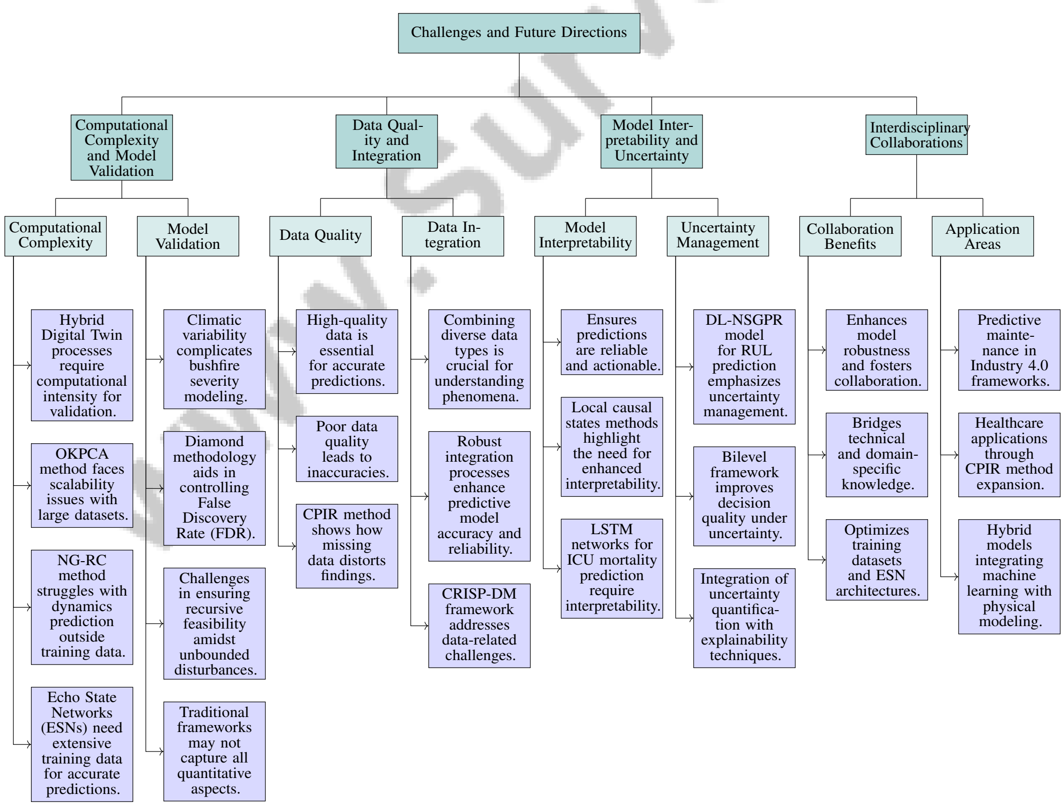


Figure 5: This figure illustrates the challenges and future directions in developing predictive models for low-probability high-impact events, focusing on computational complexity, data quality, model interpretability, and interdisciplinary collaborations.

AI-generated, for reference only.

6.1 Computational Complexity and Model Validation

Developing predictive models for low-probability high-impact events is hindered by computational complexity and model validation challenges. Large datasets and uncertainty integration increase computational demands, complicating efficient algorithm development. Hybrid Digital Twin processes exemplify the computational intensity required for validation [28]. The OKPCA method illustrates scalability issues with training dataset size, posing challenges in resource-constrained settings [34]. The NG-RC method's difficulty in predicting dynamics outside training data parameterization highlights the need for robust computational strategies [14]. Similarly, Echo State Networks (ESNs) require extensive training data coverage to predict transitions accurately, emphasizing computational burdens [51].

Model validation is complicated by factors like climatic variability in bushfire severity modeling, necessitating robust frameworks to ensure prediction reliability [12]. The Diamond methodology offers robust False Discovery Rate (FDR) control across machine learning models, aiding in discovering non-additive interactions despite data perturbations [49]. Ensuring recursive feasibility and stability amidst unbounded disturbances remains challenging, as traditional methods rely on bounded disturbances [18]. Some frameworks may not capture all quantitative aspects of high-impact events, limiting applicability [13].

Despite advancements, challenges persist in automating processes and ensuring AI-driven insights' reliability, especially in high-energy physics [59]. Future research should focus on developing validation methods for larger, diverse datasets to enhance model robustness and explore broader applicability. Addressing these challenges will better equip predictive models to manage real-world complexities and uncertainties, supporting effective risk assessment and management strategies.

6.2 Data Quality and Integration

In risk assessment for low-probability high-impact (HILP) events, data integrity and integration are crucial. HILP events involve complex threats and uncertainties requiring nuanced understanding. Effective risk management requires frameworks that balance computational efficiency with precision, necessitating high-quality data and robust integration methods [19, 13, 7, 31, 23]. High-quality data ensures accurate predictions, while poor data quality can lead to inaccuracies.

Integrating diverse data types is vital for understanding complex phenomena. For example, combining textual and numerical data is essential for accurately assessing fatalities [19]. Challenges arise when data is incomplete or of low quality, potentially skewing results. The CPIR method shows how data quality affects outcomes, as missing data can distort findings [26]. Data quality issues, such as cloud cover, can obscure vital information, affecting predictive model performance, especially with satellite-derived data [50].

Robust data integration processes are crucial for combining various data sources, enriching datasets, and enhancing predictive model accuracy and reliability. This leads to actionable insights in fields like manufacturing, data journalism, and healthcare, where systematic approaches like CRISP-DM and advanced techniques address data-related challenges [19, 25, 23]. By addressing data quality issues and ensuring effective integration, risk assessment methodologies can yield precise insights, supporting better decision-making in managing complex systems and anticipating rare events.

6.3 Model Interpretability and Uncertainty

Enhancing model interpretability and managing uncertainty are critical for predictive models addressing low-probability high-impact events. Effective interpretability ensures predictions are reliable and actionable, integrating domain knowledge with data for efficient analytics [22]. This is crucial in predictive digital twin approaches, where interpretability and uncertainty management are needed in complex systems like tumor dynamics [16].

Local causal states methods highlight the need for enhanced interpretability and uncertainty management in extreme weather events [11]. The UnTWIST method underscores challenges in naive thresholding for wave detection, emphasizing interpretability and uncertainty management [60]. The DL-NSGPR model for predicting Remaining Useful Life (RUL) underscores managing uncertainty for reliable predictions [28]. In clinical decision-making, LSTM networks predicting ICU mortality highlight the need for interpretability to support decisions [4].

AI-generated, for reference only.

Future research in Probabilistic Software Modeling (PSM) will evaluate specific applications and compare effectiveness against current methodologies, addressing interpretability challenges [61]. The bilevel framework for decision-making under uncertainty illustrates improved decision quality and effective contextual information leverage, emphasizing uncertainty management [39]. To deploy predictive models effectively, enhancing interpretability and managing uncertainty involves integrating uncertainty quantification with explainability techniques like Partial Dependence Plots (PDP) and Individual Conditional Expectation (ICE) plots, facilitating effective communication of model uncertainty. Visualization analytics enable decision-makers to derive actionable insights. A systematic approach like CRISP-DM addresses challenges in developing robust data-driven models, fostering cooperation between data and business experts [62, 23]. By overcoming these challenges, predictive models yield accurate insights, supporting decision-making and risk management in low-probability high-impact events.

6.4 Interdisciplinary Collaborations

Interdisciplinary collaborations are crucial for understanding and managing low-probability high-impact events by integrating diverse expertise and methodologies. Future research should focus on developing industry-specific extensions of the CRISP-DM framework to enhance model robustness and foster collaboration between data and business experts [23]. Such collaborations bridge technical and domain-specific knowledge, ensuring risk assessment models' effectiveness across sectors.

In predictive maintenance, interdisciplinary efforts can lead to models incorporating real-time data into Industry 4.0 frameworks, enhancing reliability and efficiency [26]. Expanding the CPIR method to include causality-driven symptom networks exemplifies interdisciplinary research potential in healthcare [26]. Optimizing training datasets and refining Echo State Network (ESN) architectures benefit from AI specialists and domain experts' collaboration, enhancing predictive capabilities across systems [51]. Optimizing the NG-RC method's scaling parameter and exploring applications in complex systems can be achieved through collaborative efforts integrating insights from multiple disciplines [14].

Interdisciplinary collaborations can enhance Case-Based Reasoning (CBR) systems' optimization algorithms, improving applicability beyond financial risk detection [15]. Developing hybrid models combining machine learning with physical modeling can enhance interpretability and address computational demands, highlighting the importance of integrating geoscience and data science expertise [33]. Interdisciplinary partnerships significantly advance predictive model development by integrating diverse expertise, enhancing accuracy, reliability, and adaptability. This approach addresses machine learning models' limitations, such as their "black box" nature and challenges in capturing complex feature interactions, as highlighted by advancements like the Diamond method for interaction discovery. Interdisciplinary efforts streamline model-driven analytics, ensuring robust insights from large datasets in healthcare and finance. By fostering cooperation between data experts and domain specialists, these partnerships overcome data-related challenges, promoting innovative applications across fields, leading to effective data-driven decision-making [22, 19, 49, 23]. Leveraging diverse expertise enhances complex systems' resilience and adaptability in low-probability high-impact events, supporting effective risk management strategies.

7 Conclusion

The survey highlights the pivotal role of merging knowledge-based and data-driven methodologies in evaluating low-probability high-impact events. This synergy not only enhances the clarity and acceptance of predictive models, as seen in clinical decision support systems for chronic diseases, but also significantly boosts predictive accuracy, exemplified by advancements in modeling the thermal behavior of composite materials. Innovative data-driven techniques, such as those achieving high classification accuracy with limited data, further demonstrate the efficiency of these integrated approaches in improving precision and resource utilization.

The urgency for robust risk management strategies is underscored by the looming threat of significant climatic shifts, necessitating immediate attention and action. Emerging probability estimation techniques reveal the underestimated importance of these events in resource allocation, prompting a reevaluation of current strategies. In the realm of epidemic control, the fusion of knowledge-based and data-driven strategies has shown promise in optimizing resource allocation, potentially curbing

AI-generated, for reference only.

the spread of infections during outbreaks. This integration is essential for developing prescriptive machine learning frameworks that support rational and ethical automated decision-making.

The survey concludes that a comprehensive methodological framework, enriched by domain expertise and ongoing refinement, is crucial for deriving meaningful insights from complex datasets. By leveraging these integrated methodologies, decision-makers can effectively navigate the uncertainties inherent in complex systems, thereby enhancing policy development and strategic planning. The qualitative framework for analyzing high-impact low-probability events serves as a crucial tool for mitigating uncertainties and refining decision-making processes.

AI-generated, for reference only.

References

[1] Peter D. Ditlevsen and Susanne Ditlevsen. Warning of a forthcoming collapse of the atlantic meridional overturning circulation, 2023.

[2] Alyssa Cassity, Hieu Le, Hernan Santos, Erik Priest, and Jian Tao. Dynamic data-driven digital twin testbed for enhanced first responder training and communication, 2024.

[3] Kayvan Miri Lavassani, Zachary M. Boyd, Bahar Movahedi, and Jason Vasquez. Ten-tier and multi-scale supplychain network analysis of medical equipment: Random failure and intelligent attack analysis, 2023.

[4] Manel Mili, Asma Kerkeni, Asma Ben Abdallah, and Mohamed Hedi Bedoui. Icu mortality prediction using long short-term memory networks, 2023.

[5] Aaron Clauset and Ryan Woodard. Estimating the historical and future probabilities of large terrorist events, 2014.

[6] David Scott Hunter, Ajay Saini, and Tauhid Zaman. Picking winners: A data driven approach to evaluating the quality of startup companies, 2018.

[7] Iver Bakken Sperstad and Erlend Sandø Kiel. Development of a qualitative framework for analysing high-impact low-probability events in power systems. In *Safety and Reliability–Safe Societies in a Changing World*, pages 1599–1607. CRC Press, 2018.

[8] Sakshi Mishra, Ted Kwasnik, and Kate Anderson. Microgrid resilience: A holistic and context-aware resilience metric, 2021.

[9] Yang Li. Enhancing resilience of power systems against typhoon threats: A hybrid data-model driven approach, 2024.

[10] Daniel Bienstock, Yury Dvorkin, Cheng Guo, Robert Mieth, and Jiayi Wang. Risk-aware security-constrained unit commitment: Taming the curse of real-time volatility and consumer exposure, 2024.

[11] Adam Rupe, Karthik Kashinath, Nalini Kumar, and James P. Crutchfield. Unsupervised discovery of extreme weather events using universal representations of emergent organization, 2023.

[12] Shouthiri Partheepan, Farzad Sanati, and Jahan Hassan. Bushfire severity modelling and future trend prediction across australia: Integrating remote sensing and machine learning, 2024.

[13] Aaron Velasco, Olga Kosheleva, and Vladik Kreinovich. Low-probability high-impact events are even more important than it is usually assumed. 2023.

[14] Daniel Köglmayr and Christoph Räth. Extrapolating tipping points and simulating non-stationary dynamics of complex systems using efficient machine learning, 2023.

[15] Wei Li, Florentina Paraschiv, and Georgios Sermpinis. A data-driven explainable case-based reasoning approach for financial risk detection, 2021.

[16] Mohammad Azangoo, Joonas Salmi, Iivo Yrjölä, Jonathan Bensky, Gerardo Santillan, Nikolaos Papakonstantinou, Seppo Sierla, and Valeriy Vyatkin. Hybrid digital twin for process industry using apros simulation environment, 2021.

[17] Julian Berberich and Frank Allgöwer. An overview of systems-theoretic guarantees in data-driven model predictive control, 2024.

[18] Guanru Pan, Ruchuan Ou, and Timm Faulwasser. Towards data-driven stochastic predictive control, 2022.

[19] Thomas Rusch, Paul Hofmarcher, Reinhold Hatzinger, and Kurt Hornik. Model trees with topic model preprocessing: An approach for data journalism illustrated with the wikileaks afghanistan war logs, 2013.

[20] Emanuele Mengoli, Zhiyuan Yao, and Wutao Wei. Develop end-to-end anomaly detection system, 2024.

[21] Eyke Hüllermeier. Prescriptive machine learning for automated decision making: Challenges and opportunities, 2021.

[22] Thomas Hartmann, Assaad Moawad, Francois Fouquet, Gregory Nain, Jacques Klein, Yves Le Traon, and Jean-Marc Jezequel. Model-driven analytics: Connecting data, domain knowledge, and learning, 2017.

[23] Shailesh Tripathi, David Muhr, Brunner Manuel, Frank Emmert-Streib, Herbert Jodlbauer, and Matthias Dehmer. Ensuring the robustness and reliability of data-driven knowledge discovery models in production and manufacturing, 2020.

[24] Kristen A Severson, Peter M Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H Chen, Muratahan Aykol, Patrick K Herring, Dimitrios Fraggedakis, et al. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5):383–391, 2019.

[25] Jonathan H Chen and Steven M Asch. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *The New England journal of medicine*, 376(26):2507, 2017.

[26] Abd AlRahman AlMomani and Erik Bollt. Informative ranking of stand out collections of symptoms: A new data-driven approach to identify the strong warning signs of covid 19, 2020.

[27] Jiajia Li, Feng Tan, Cheng He, Zikai Wang, Haitao Song, Lingfei Wu, and Pengwei Hu. Higenet: A highly efficient modeling for long sequence time series prediction in aiops, 2022.

[28] Zhaoyi Xu, Yanjie Guo, and Joseph Homer Saleh. Accurate remaining useful life prediction with uncertainty quantification: a deep learning and nonstationary gaussian process approach, 2021.

[29] Ceyda Yaba Best, Amin Khademi, and Burak Eksioglu. Data-driven infectious disease control with uncertain resources, 2020.

[30] William Schueller, Christian Diem, Melanie Hinterplattner, Johannes Stangl, Beate Conrady, Markus Gerschberger, and Stefan Thurner. Propagation of disruptions in supply networks of essential goods: A population-centered perspective of systemic risk, 2022.

[31] Joakim Sundh. Human behavior in the context of low-probability high-impact events. *Humanities and Social Sciences Communications*, 11(1):1–10, 2024.

[32] Zara Lisbon. Review of digital asset development with graph neural network unlearning, 2024.

[33] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and F Prabhat. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, 2019.

[34] Zachary Morrison, Benjamin P. Russo, Yingzhao Lian, and Rushikesh Kamalapurkar. Fault detection via occupation kernel principal component analysis, 2023.

[35] Gregory G. Garner and Klaus Keller. When tails wag the decision: The role of distributional tails on climate impacts on decision-relevant time-scales, 2017.

[36] Simon H. Tindemans and Goran Strbac. Robust estimation of risks from small samples, 2017.

[37] Dario Piga, Marco Forgione, Simone Formentin, and Alberto Bemporad. Performance-oriented model learning for data-driven mpc design, 2019.

[38] Atul Agrawal and Phaedon-Stelios Koutsourelakis. A probabilistic, data-driven closure model for rans simulations with aleatoric, model uncertainty, 2024.

[39] Miguel Angel Muñoz, Salvador Pineda, and Juan Miguel Morales. A bilevel framework for decision-making under uncertainty with contextual information, 2021.

[40] Kevin Zhang and Hod Lipson. Aligning ai-driven discovery with human intuition, 2024.

AI-generated, for reference only.

[41] Christopher T. Franck and Christopher E. Wilson. Predicting competitions by combining conditional logistic regression and subjective bayes: An academy awards case study, 2024.

[42] Xin Wang, Wei Xue, Yilun Han, and Guangwen Yang. Efficient climate simulation via machine learning method, 2022.

[43] Kevin Schultz, Marisel Villafane-Delgado, Elizabeth P. Reilly, Grace M. Hwang, and Anshu Saksena. Graph signal processing for infrastructure resilience: Suitability and future directions, 2020.

[44] Tao Han, Zhenghao Chen, Song Guo, Wanghan Xu, and Lei Bai. Cra5: Extreme compression of era5 for portable global climate and weather research via an efficient variational transformer, 2024.

[45] Namyong Park, MinHyeok Kim, Nguyen Xuan Hoai, R. I., McKay, and Dong-Kyun Kim. Knowledge-guided dynamic systems modeling: A case study on modeling river water quality, 2021.

[46] Tian Tian and Jiahao Deng. Unleashing the power of ai: Transforming marketing decision-making in heavy machinery with machine learning, radar chart simulation, and markov chain analysis, 2024.

[47] Tito Homem de Mello, Juan Valencia, Felipe Lagos, and Guido Lagos. Forecasting outside the box: Application-driven optimal pointwise forecasts for stochastic optimization, 2024.

[48] Rohit Kannan, Güzin Bayraksan, and James R. Luedtke. Data-driven sample average approximation with covariate information, 2022.

[49] Winston Chen, Yifan Jiang, William Stafford Noble, and Yang Young Lu. Error-controlled non-additive interaction discovery in machine learning models, 2024.

[50] Srija Chakraborty and Eleanor C. Stokes. Adaptive modeling of satellite-derived nighttime lights time-series for tracking urban change processes using machine learning, 2023.

[51] Anton Pershin, Cedric Beaume, Kuan Li, and Steven M. Tobias. Can neural networks predict dynamics they have never seen?, 2021.

[52] Yinpu Li, Siqi Mao, Yaping Yuan, Ziren Wang, Yixin Kang, and Yuanxin Yao. Beyond tides and time: Machine learning triumph in water quality, 2023.

[53] Tyrus Berry and John Harlim. Semiparametric forecasting and filtering: correcting low-dimensional model error in parametric models, 2015.

[54] Francesco Borra, Angelo Vulpiani, and Massimo Cencini. Effective models and predictability of chaotic multiscale systems via machine learning, 2020.

[55] Jessica Leoni, Valentina Breschi, Simone Formentin, and Mara Tanelli. Explainable data-driven modeling via mixture of experts: towards effective blending of grey and black-box models, 2024.

[56] Syed Zahid Husain, Leo Separovic, Jean-François Caron, Rabah Aider, Mark Buehner, Stéphane Chamberland, Ervig Lapalme, Ron McTaggart-Cowan, Christopher Subich, Paul A. Vaillancourt, Jing Yang, and Ayrton Zadra. Leveraging data-driven weather models for improving numerical weather prediction skill through large-scale spectral nudging, 2024.

[57] Christopher Bülte, Nina Horat, Julian Quinting, and Sebastian Lerch. Uncertainty quantification for data-driven weather models, 2024.

[58] Vignesh Gopakumar, Ander Gray, Joel Oskarsson, Lorenzo Zanisi, Stanislas Pamela, Daniel Giles, Matt Kusner, and Marc Peter Deisenroth. Uncertainty quantification of surrogate models using conformal prediction, 2024.

[59] Peter W. Hatfield, Jim A. Gaffney, Gemma J. Anderson, Suzanne Ali, Luca Antonelli, Suzan Başeğmez du Pree, Jonathan Citrin, Marta Fajardo, Patrick Knapp, Brendan Kettle, Bogdan Kustowski, Michael J. MacDonald, Derek Mariscal, Madison E. Martin, Taisuke Nagayama, Charlotte A. J. Palmer, J. Luc Peterson, Steven Rose, J J Ruby, Carl Shneider, Matt J. V. Streeter, Will Trickey, and Ben Williams. The data-driven future of high energy density physics, 2021.

[60] Ariana Mendible, Weston Lowrie, Steven L. Brunton, and J. Nathan Kutz. Data-driven modeling of two-dimensional detonation wave fronts, 2021.

[61] Hannes Thaller, Lukas Linsbauer, Rudolf Ramler, and Alexander Egyed. Probabilistic software modeling: A data-driven paradigm for software analysis, 2019.

[62] Nijat Mehdiyev, Maxim Majlatow, and Peter Fettke. Communicating uncertainty in machine learning explanations: A visualization analytics approach for predictive process monitoring, 2023.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.