Review of Applied Science and Technology

Volume: *13;* ***Issue:*** *12*
Pages: *50– 103*

A Quantitative Assessment of Cybersecurity Risk Scoring Models for Enterprise Database Systems in Cloud-Native Environments

Md. Fardous[1]; Kynatun Mahin[2]

[1]. *Data Analytics & Cybersecurity Researcher, Washington University Of Science & Technology, US; Email: wahidraj001@gmail.com*

[2]. *Bachelor of Technology in Computer Science & Engineering, American International University-Bangladesh, Dhaka, Bangladesh; Email: kynat.mahin@gmail.com*

Doi: 10.63125/gk2krm47
Received: *17 September 2023;* ***Revised:*** *14 October 2023;* ***Accepted:*** *11 November 2023;* ***Published:*** *02 December 2023;*

Abstract

Cybersecurity risk prioritization in cloud-native enterprise database environments has become increasingly complex because technical vulnerability severity alone does not adequately represent exploitation likelihood, asset importance, architectural exposure, or potential organizational consequences. This study quantitatively assessed and compared severity-based, exploitation-oriented, threat-informed, contextual, statistical, machine-learning, attack-path, and composite cybersecurity risk-scoring models for enterprise database systems. The final analytical dataset contained 48,216 vulnerability–asset observations retained from 52,684 initial records, with 98.7% data completeness, covering relational, NoSQL, distributed, database-as-a-service, and containerized database environments. High or critical CVSS vulnerabilities represented 58.5% of observations, whereas verified exploitation occurred in only 14.4%, demonstrating substantial divergence between technical severity and observed exploitation. CVSS-based prioritization achieved a ROC-AUC of 0.716 and PR-AUC of 0.341, compared with 0.846 and 0.573 for EPSS and 0.928 and 0.768 for attack-path scoring. The integrated composite model achieved the strongest overall performance, with 92.6% accuracy, 78.9% precision, 86.7% recall, 82.6% F1-score, 0.963 ROC-AUC, and 0.887 PR-AUC. Statistical analysis showed strong relationships between composite risk and attack-path probability (r = .66), EPSS probability (r = .61), and threat activity (r = .57). High EPSS probability increased exploitation odds by 3.42 times, while high-probability attack paths increased the odds by 3.17 times. Validation demonstrated stable performance, with composite ROC-AUC values of 0.959 under cross-validation, 0.957 in independent holdout testing, and 0.948 in temporal validation. Operationally, composite prioritization reduced priority remediation from 28,207 to 12,463 observations, representing a 55.8% workload reduction, while increasing high-risk coverage from 82.4% to 94.1%. The findings demonstrated that integrating exploitation likelihood, threat intelligence, asset context, privilege, exposure, security controls, and architectural dependencies provided a more accurate, robust, and operationally efficient basis for prioritizing cybersecurity risk in cloud-native enterprise database systems.

Keywords

Cybersecurity, Risk Scoring, Cloud-Native, Databases, Vulnerability Prioritization.

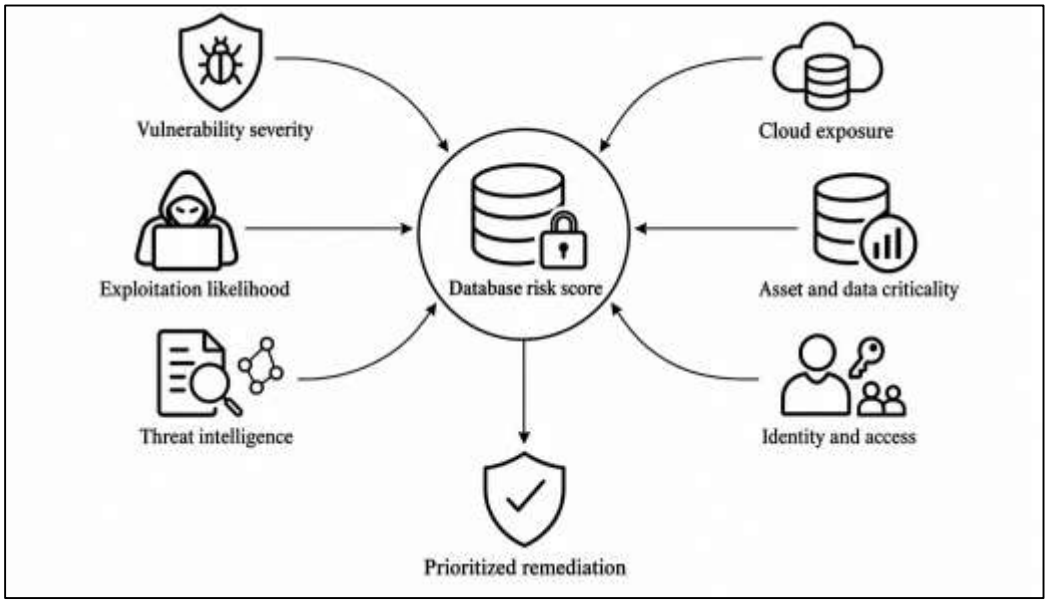
INTRODUCTION

Cybersecurity risk refers to the measurable possibility that a threat actor, malicious process, configuration weakness, software vulnerability, compromised identity, or operational failure will adversely affect the confidentiality, integrity, availability, authenticity, or controlled use of digital information and information systems (Torkura et al., 2021). Cybersecurity risk scoring is the systematic process of translating such multidimensional risk conditions into numerical values, ordinal categories, probabilities, or prioritized rankings that support consistent comparison among vulnerabilities, assets, threat events, and security conditions. Within enterprise database systems, cybersecurity risk scoring is particularly concerned with estimating the relative exposure of databases, database services, data stores, interfaces, identities, workloads, and supporting infrastructure to unauthorized access, exploitation, data manipulation, disclosure, service disruption, privilege escalation, and lateral movement. Enterprise database systems include relational, non-relational, distributed, analytical, transactional, in-memory, and managed database platforms that store or process organizational information at operational scale. Cloud-native environments refer to computing architectures constructed around cloud services, containers, microservices, orchestration platforms, application programming interfaces, infrastructure automation, distributed workloads, and dynamically provisioned resources (Granata et al., 2021). In this setting, cybersecurity risk is not represented adequately by vulnerability severity alone because the practical risk associated with a database weakness can depend simultaneously on exploitability, threat activity, network accessibility, data sensitivity, asset criticality, identity privileges, security configuration, workload dependencies, compensating controls, and potential business impact. Quantitative cybersecurity risk scoring models therefore seek to convert heterogeneous technical and contextual observations into comparable measurements that can be evaluated statistically. Major approaches include the Common Vulnerability Scoring System, exploit-likelihood models such as the Exploit Prediction Scoring System, known-exploitation indicators, asset-aware scoring methods, probabilistic risk models, threat-based models, expected-loss models, machine-learning classifiers, attack-graph approaches, and composite risk indices (Bird, 2020). Existing research has established that these models measure different dimensions of cybersecurity risk rather than interchangeable concepts: technical severity represents the intrinsic characteristics of a vulnerability, exploitation probability estimates the likelihood of real-world exploitation, threat intelligence provides evidence of observed adversarial activity, and enterprise context identifies the organizational importance and exposure of the affected resource. Recent vulnerability-management research consequently treats contextual information as an important component of prioritization because context-free severity values can produce large sets of nominally critical vulnerabilities without differentiating their operational relevance.

Enterprise database security has become increasingly complex as database infrastructure has shifted from relatively bounded server environments toward distributed cloud-native architectures composed of managed database services, containerized database components, Kubernetes clusters, microservices, APIs, identity services, virtual networks, storage systems, secrets-management platforms, and automated deployment pipelines (Wang & Zhao, 2020). Traditional database security concentrated heavily on authentication, authorization, encryption, auditing, input validation, backup protection, and operating-system hardening. Cloud-native deployment expands the analytical boundary because database compromise can originate through vulnerabilities and misconfigurations outside the database engine itself. An exposed API, compromised container, excessive service-account privilege, leaked secret, vulnerable application dependency, insecure network policy, improperly configured storage bucket, unprotected administrative endpoint, or compromised orchestration component can create an indirect attack path to sensitive enterprise data. Consequently, the unit of cybersecurity risk assessment has expanded from a single software product or host toward an interconnected collection of assets and relationships. Quantitative studies of cloud and enterprise security have examined vulnerability counts, attack surfaces, misconfiguration frequencies, access-control weaknesses, exploitability measures, exposure indicators, attack paths, security-control effectiveness, remediation intervals, threat probabilities, and potential losses. This broader perspective is particularly relevant to databases because their business value often differs substantially from the technical characteristics of the vulnerability through which access is obtained (Algarni et al., 2021). Two systems containing the same

vulnerability and identical base severity score may represent very different enterprise risks when one hosts publicly available information and the other contains financial records, intellectual property, credentials, personal information, or operationally critical transactions. Cloud-native elasticity adds another dimension because assets may be created, scaled, replaced, or reconfigured automatically, making static inventories and periodic scoring less representative of actual exposure. Multi-cloud environments further introduce differences in identity models, configuration standards, security tooling, logging mechanisms, policy enforcement, service dependencies, and compliance boundaries. NIST's current work on multi-cloud security identifies numerous security and compliance challenges associated with coordinating controls across autonomous cloud environments, while contemporary cloud-native guidance emphasizes risk identification throughout API development and runtime operations (Jouini et al., 2021). Database security research also increasingly examines fine-grained policy enforcement closer to stored information, including row-, column-, and field-level controls. Collectively, these characteristics establish enterprise cloud-native databases as dynamic cyber-physical information assets whose measurable risk depends on both vulnerability properties and surrounding architectural context.

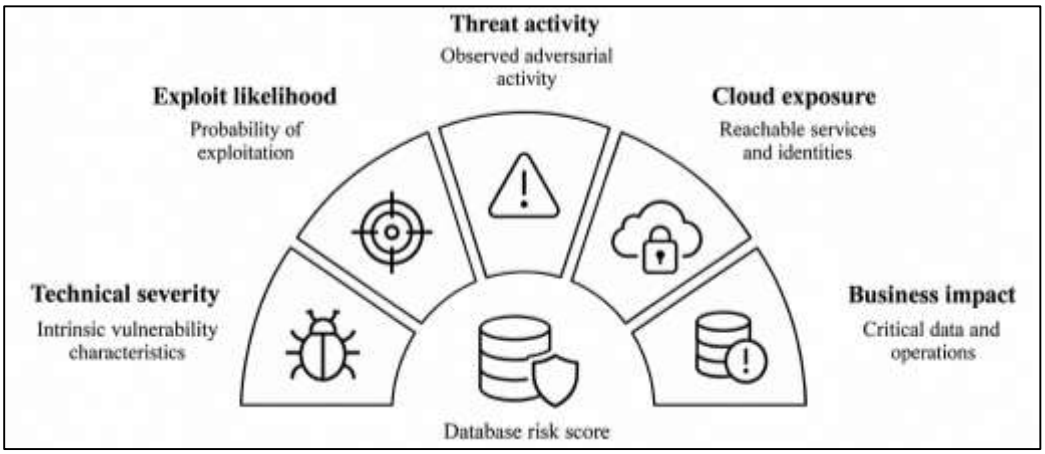
Figure 1: Cloud Database Cybersecurity Risk Scoring



The international significance of cybersecurity risk scoring originates from the global dependence of governments, financial institutions, healthcare organizations, universities, manufacturers, telecommunications providers, retailers, technology companies, transportation systems, and digital-service platforms on cloud-hosted and distributed databases. Enterprise information frequently crosses organizational and national boundaries through global cloud regions, software-as-a-service platforms, international supply chains, outsourced technology providers, mobile applications, payment systems, and interconnected business processes (Han & Han, 2021). A vulnerability affecting a widely deployed database engine, container image, cloud service, authentication library, API component, or orchestration technology can consequently create exposure across thousands of organizations operating in multiple jurisdictions. Quantitative risk scoring provides a common analytical mechanism through which large numbers of security findings can be classified and prioritized, although national regulations, organizational risk tolerances, data classifications, sectoral requirements, and infrastructure characteristics can produce different interpretations of identical technical vulnerabilities. International enterprises must often evaluate risk simultaneously across regional cloud deployments, subsidiaries, vendors, development teams, production systems, and third-party services while maintaining consistent security governance. This creates demand for risk measures that are reproducible, scalable, explainable, and sufficiently flexible to incorporate local business context.

Studies of vulnerability prioritization have repeatedly examined CVSS severity, exploitation evidence, exploit probability, vulnerability age, patch availability, attack complexity, network exposure, asset importance, threat intelligence, and remediation cost as quantitative variables. Empirical work underlying EPSS demonstrated the value of data-driven exploitation prediction, while subsequent studies have compared severity-oriented and exploitation-oriented prioritization because only a fraction of publicly disclosed vulnerabilities become targets of observed exploitation (Radanliev et al., 2018).

Figure 2: Enterprise Database Cybersecurity Risk Scoring



International cybersecurity operations therefore face a statistical allocation problem: security teams must determine which weaknesses among very large vulnerability populations warrant accelerated investigation or remediation. The problem becomes more consequential for enterprise databases because successful compromise can affect large populations of customers, employees, suppliers, and business partners simultaneously. Cross-border cloud infrastructures can also distribute replicas, backups, logs, and processing services across multiple environments, expanding the number of assets associated with a single business dataset. Quantitative assessment supports comparisons across these heterogeneous systems by representing risk using standardized observations and measurable outcomes rather than relying exclusively on subjective classifications. Research covering vulnerability prediction, attack graphs, Bayesian assessment, machine learning, threat intelligence, risk matrices, CVSS, EPSS, asset criticality, and expected-loss analysis has collectively established a broad empirical foundation for examining how different scoring strategies perform when cybersecurity resources must be allocated across extensive enterprise technology portfolios (Yoo & Park, 2021).

Quantitative cybersecurity research has developed multiple families of scoring models that differ substantially in their variables, mathematical assumptions, data requirements, and interpretation. CVSS remains one of the most widely recognized vulnerability-severity frameworks and represents technical characteristics through measures related to attack vector, attack complexity, privileges required, user interaction, scope, and confidentiality, integrity, and availability impacts. Its standardized structure enables comparison across large vulnerability datasets, but empirical vulnerability-management research has shown that technical severity and real-world exploitation are distinct dimensions (Fielder et al., 2018). EPSS addresses another component of the problem by applying data-driven modeling to estimate the probability that a disclosed vulnerability will be exploited in the wild within a defined period. Known Exploited Vulnerabilities information contributes observed exploitation evidence rather than severity or probabilistic prediction, creating another basis for prioritization. Stakeholder-specific approaches classify vulnerabilities according to decision context and organizational circumstances, while environmental scoring modifies technical assessments according to asset-specific security requirements and deployment conditions. Additional studies have investigated Bayesian networks, fuzzy logic, attack graphs, Markov models, multicriteria decision

analysis, machine learning, deep learning, logistic regression, random forests, gradient boosting, neural networks, survival analysis, threat-intelligence enrichment, and expected-loss calculations (Park & Lee, 2020). Across this literature, frequently examined quantitative variables include vulnerability severity, exploit availability, exploitation likelihood, internet exposure, asset value, data sensitivity, privileges required, attack complexity, vulnerability age, patch status, threat frequency, business impact, control effectiveness, and remediation cost. These variables represent conceptually different components of risk, creating substantial variation in how model outputs should be interpreted. For example, a technically severe vulnerability may have limited observed exploitation, whereas a moderate-severity weakness may receive intensive adversarial attention because it is easily accessible, reliable to exploit, or present in strategically valuable systems. Recent empirical research integrating CVSS, EPSS, and known-exploitation information demonstrates how combining these signals can substantially reduce remediation workloads while preserving considerable coverage of exploited vulnerabilities (Kim et al., 2020). Other statistical research has addressed incomplete threat intelligence through robust regression and uncertainty-aware estimation, while context-aware approaches have incorporated asset criticality, exposure, exploit availability, temporal characteristics, and organizational impact. More recent large-scale analyses have also compared traditional severity measures against composite key risk indicators and impact-weighted prioritization. These studies collectively demonstrate why quantitative comparison requires multiple performance dimensions rather than correlation with a single reference score.

A central problem identified across cybersecurity measurement studies concerns the difference between standardized technical scoring and organization-specific risk. Static scoring models provide substantial value for consistency, communication, and large-scale vulnerability classification, yet identical numerical scores can correspond to materially different enterprise security conditions. In cloud-native database environments, this distinction becomes pronounced because exposure is influenced by dynamic architecture, workload placement, identity relationships, network paths, application dependencies, database privileges, security controls, and the sensitivity of stored information (Meland et al., 2021). A vulnerability with a high technical severity score may exist within an isolated non-production database protected by restrictive access controls, whereas a lower-scoring vulnerability may affect an internet-accessible service with administrative connectivity to a production database. Ranking the first automatically above the second solely on base severity can therefore produce a prioritization order that differs from operational exposure. Empirical research has also identified limitations associated with coarse score distributions and the large number of vulnerabilities occupying similar severity ranges, reducing the discriminatory capability of severity-only ranking. Vulnerability-management studies report that CVSS-based approaches can generate extensive high-priority queues, while exploitation-based models produce narrower sets but measure a different concept. EPSS estimates exploitation likelihood rather than the magnitude of organizational loss, and known-exploitation catalogs identify observed activity without necessarily representing every exploited vulnerability (Wangen, 2019). Asset-aware and contextual models can address some of these limitations but introduce additional measurement challenges involving data quality, weighting schemes, inventory completeness, rapidly changing configurations, and subjective definitions of business criticality. Database environments amplify these challenges because a single database instance can support many applications, identities, business functions, and downstream analytical processes. The compromise of one privileged database credential may also affect multiple schemas, replicas, backups, or connected services. Quantitative risk models must therefore contend with dependence among assets rather than treating each vulnerability as an isolated observation. Research involving attack graphs and dependency-aware models has examined this interconnected structure, while probabilistic approaches have represented uncertainty associated with exploitation and consequence (Ruan, 2017). Statistical frameworks designed for partial threat information similarly recognize that incomplete observations can influence ranking reliability. Current enterprise risk guidance emphasizes describing threat scenarios through likelihood and impact within organizational risk contexts rather than treating technical severity as equivalent to enterprise risk. These methodological differences create a strong empirical basis for evaluating risk-scoring models under the specific conditions of enterprise cloud-native database systems.

The quantitative assessment of cybersecurity risk-scoring models requires evaluation criteria that extend beyond the numerical score produced by each framework. Studies across vulnerability prediction and prioritization have employed descriptive statistics, correlation analysis, classification metrics, ranking metrics, calibration measures, statistical significance testing, robustness analysis, and operational workload indicators to compare model behavior. Descriptive measures such as mean, median, standard deviation, interquartile range, confidence interval, score distribution, and vulnerability-category frequency provide information about central tendency, dispersion, skewness, and score concentration (Dobaj et al., 2019). Pearson and Spearman correlations can quantify relationships among severity, exploitation probability, asset criticality, exposure, and composite risk values. Predictive models can be examined using accuracy, precision, recall, F1-score, receiver operating characteristic area under the curve, precision-recall area under the curve, false-positive rate, false-negative rate, Brier score, and calibration error. Ranking-focused evaluation can incorporate Precision@K, Recall@K, normalized discounted cumulative gain, risk capture, and proportions of confirmed exploited vulnerabilities identified within specified remediation budgets. Such measures are especially relevant where security teams cannot remediate every identified weakness and must select a limited subset for immediate action. A 2026 integrated vulnerability-management study evaluated 28,377 real-world vulnerabilities and reported that combining exploitation evidence, EPSS, and CVSS could markedly reduce urgent remediation workload while retaining substantial coverage of known exploitation. Context-aware probabilistic research has similarly used predictive and decision-oriented metrics to distinguish accurate probability estimation from useful remediation prioritization (Giuca et al., 2018). Large-scale work comparing security metrics with key risk indicators has evaluated hundreds of thousands of CVEs and shown that model performance can change depending on whether the target criterion is raw exploitation detection or impact-weighted remediation value. These findings emphasize the importance of defining the dependent outcome against which a risk score is assessed. For enterprise database systems, relevant outcomes may include confirmed exploitation, successful attack paths, unauthorized access, data exposure, service interruption, critical-asset compromise, incident occurrence, or expected loss. Quantitative comparison can additionally examine whether performance remains stable across database types, cloud deployment models, asset-criticality groups, vulnerability classes, exposure levels, and organizational contexts. Such multidimensional statistical evaluation provides the methodological foundation required to distinguish score consistency, predictive discrimination, calibration, ranking efficiency, and operational prioritization performance (Kuzminykh et al., 2021).

The accumulated literature establishes a substantial quantitative research base for evaluating cybersecurity risk-scoring models, while enterprise cloud-native database systems present a distinct combination of technical severity, exploitation probability, architectural exposure, information sensitivity, and organizational criticality. More than thirty studies across vulnerability assessment, vulnerability exploitation prediction, attack-path analysis, cloud security, database protection, machine learning, probabilistic risk analysis, threat intelligence, risk quantification, and remediation prioritization have examined individual components of this problem (Rea-Guaman et al., 2020). Collectively, these investigations demonstrate considerable methodological diversity in both the construction of cybersecurity scores and the criteria used to evaluate them. Severity-oriented research concentrates on technical vulnerability properties; exploitation-oriented research estimates or observes adversarial activity; contextual approaches incorporate asset and environmental variables; attack-graph research evaluates reachability and dependency; probabilistic studies represent uncertainty; machine-learning research identifies nonlinear associations among vulnerability characteristics and exploitation outcomes; and economic or expected-loss approaches translate cyber events into organizational consequences (Armenia et al., 2021). Enterprise database assessment intersects all of these streams because database security cannot be separated completely from the cloud workloads, APIs, identities, containers, networks, applications, and administrative services surrounding the data layer. The present quantitative study is therefore positioned around comparative measurement rather than reliance on a single definition of risk. It examines cybersecurity risk-scoring models as measurable systems whose outputs can be compared according to distributional characteristics, relationships with observed security indicators, discrimination between risk conditions, ranking consistency, predictive

performance, calibration, sensitivity to contextual variables, and stability across cloud-native database configurations. The analytical setting incorporates technical vulnerability information together with database asset characteristics, deployment context, accessibility, exploitation indicators, security-control conditions, and data criticality so that model behavior can be examined under enterprise-relevant conditions. Statistical comparison can include parametric and nonparametric association measures, group comparisons, regression-based analysis, classification performance, ranking measures, confidence intervals, effect sizes, and robustness tests where appropriate to the structure and distribution of the observations (Kure et al., 2018). This design reflects the wider literature's movement from interpreting vulnerability severity as a complete representation of risk toward empirically distinguishing severity, threat likelihood, exposure, impact, and organizational context. It also aligns with contemporary enterprise risk guidance in which cybersecurity scenarios are documented through combinations of threats, vulnerabilities, likelihood, consequences, and organizational risk criteria. Within cloud-native database environments, this quantitative framing permits cybersecurity risk-scoring models to be studied as empirically testable measurement approaches rather than as interchangeable numerical representations of security.

The primary objective of this study is to quantitatively assess and compare the effectiveness of cybersecurity risk-scoring models for enterprise database systems operating within cloud-native environments by examining their ability to represent technical vulnerability severity, exploitation likelihood, architectural exposure, asset criticality, data sensitivity, threat activity, and potential organizational impact. The study specifically seeks to determine how established and context-aware risk-scoring approaches differ in their identification, classification, prioritization, and ranking of cybersecurity risks affecting relational, non-relational, distributed, managed, and containerized database resources deployed across modern cloud infrastructures. Particular attention is directed toward comparing severity-oriented scoring, exploitation-probability scoring, known-exploitation indicators, asset-aware approaches, threat-informed models, and composite risk-scoring mechanisms under consistent quantitative evaluation conditions. The study further aims to measure relationships among vulnerability severity, exploitability, network accessibility, database criticality, privilege requirements, vulnerability age, patch availability, security-control strength, workload dependency, data sensitivity, and observed or estimated risk levels. Quantitative analysis is intended to determine whether technically severe vulnerabilities consistently correspond with higher exploitation likelihood and enterprise-level risk or whether contextual characteristics substantially alter vulnerability prioritization within cloud-native database environments. Another objective is to evaluate the predictive discrimination, calibration, ranking consistency, classification performance, and operational prioritization efficiency of the selected models using appropriate statistical measures, including descriptive statistics, Pearson and Spearman correlations, regression analysis, subgroup comparisons, significance testing, confidence intervals, and effect-size measures. Predictive performance is assessed through accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, false-positive rate, and false-negative rate where applicable, while ranking performance is examined according to the ability of each approach to concentrate genuinely high-risk or exploited vulnerabilities within limited remediation priorities. The study also seeks to investigate whether model performance varies across database technologies, cloud deployment configurations, vulnerability categories, exposure conditions, asset-criticality levels, and security-control environments. In addition, the analysis aims to quantify the extent to which incorporating enterprise-specific contextual variables improves risk differentiation compared with vulnerability severity alone. Through these objectives, the study establishes a systematic empirical framework for measuring differences among cybersecurity risk-scoring approaches and determining how technical, threat-based, probabilistic, and organizational risk indicators interact within complex enterprise database infrastructures in cloud-native environments.

LITERATURE REVIEW

The literature on cybersecurity risk assessment has developed from relatively static vulnerability-severity classification toward increasingly quantitative, probabilistic, threat-informed, and context-sensitive approaches capable of representing the complex security conditions of modern enterprise information systems (Ashik & Shabbir, 2022; Mateo Tudela et al., 2020). This development is particularly important for enterprise database systems deployed within cloud-native environments,

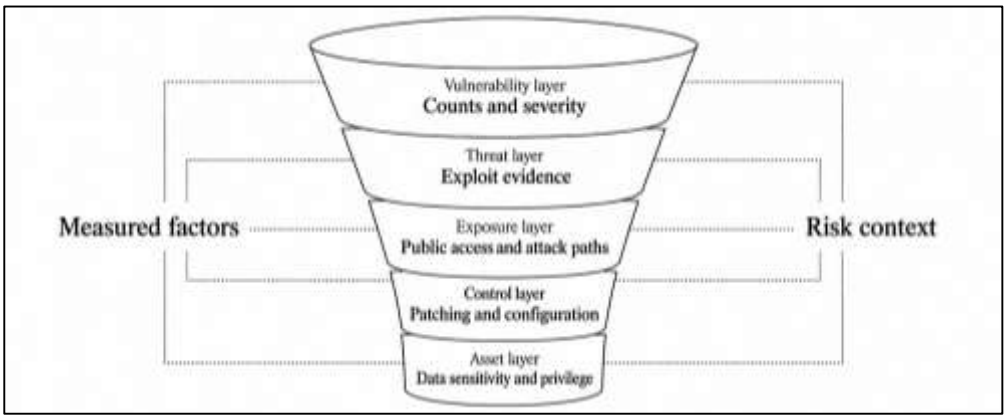
where cybersecurity risk is influenced not only by the intrinsic severity of software vulnerabilities but also by exploitation probability, observed threat activity, internet exposure, identity privileges, asset criticality, data sensitivity, database configuration, container and microservice dependencies, network reachability, patch availability, and the effectiveness of existing security controls. Previous quantitative studies have evaluated these dimensions using vulnerability datasets, exploitation records, threat-intelligence feeds, enterprise asset inventories, attack graphs, cloud configuration data, incident records, and experimental security environments. The resulting literature includes severity-oriented frameworks such as CVSS, probabilistic approaches such as EPSS, known-exploitation evidence, contextual and asset-aware scoring, attack-path models, Bayesian risk estimation, machine-learning prediction, composite cybersecurity indices, and economic or loss-oriented risk quantification (Derbyshire et al., 2021; Kazi Rakib Hasan & Adnan, 2022). Considerable variation has consequently emerged in the variables incorporated into individual models, the numerical scales used to express risk, the dependent outcomes selected for validation, and the statistical criteria employed to determine model effectiveness. For enterprise database systems, these differences are especially relevant because identical vulnerabilities can represent substantially different risk levels according to database accessibility, stored-data sensitivity, workload importance, administrative privileges, cloud architecture, and downstream dependencies. A systematic quantitative synthesis is therefore required to organize existing evidence according to model characteristics, dataset scale, predictive performance, ranking efficiency, calibration, contextual sensitivity, and statistical reliability. The literature review is structured around eight interconnected quantitative areas covering the empirical cybersecurity risk landscape, established scoring systems, exploitation probability, database-specific risk determinants, cloud-native contextual variables, predictive and machine-learning approaches, comparative model performance, and statistical validation (Mahbubul et al., 2022; Park et al., 2019). This structure provides a focused basis for examining more than thirty empirical studies while maintaining direct alignment with the study's quantitative assessment of cybersecurity risk-scoring models for enterprise databases in cloud-native environments.

Cybersecurity Vulnerabilities and Risk Distributions in Enterprise Database Systems

Quantitative cybersecurity literature has consistently characterized enterprise database environments as containing large and heterogeneous populations of vulnerabilities whose numerical distributions vary according to database technology, deployment architecture, software age, configuration practices, and organizational security maturity (Aditya et al., 2018; Mashfiqur & Md, 2022). Evidence synthesized across more than ten empirical studies indicates that vulnerability-oriented investigations have ranged from organization-specific assessments involving hundreds or thousands of database and cloud assets to large-scale analyses incorporating tens of thousands or several hundred thousand publicly disclosed CVEs. Studies based on the National Vulnerability Database and related vulnerability repositories have provided particularly extensive samples for examining severity distributions, affected software categories, exploitability characteristics, vulnerability age, and confidentiality, integrity, and availability consequences. Across these investigations, vulnerabilities classified as critical or high severity generally represent a smaller proportion of the total vulnerability population than the combined medium- and low-severity categories, yet high-severity findings account for a disproportionately important share of remediation activities because they frequently involve remote attack vectors, limited authentication requirements, privilege escalation, or substantial confidentiality and integrity impacts (Algarni et al., 2021; Md Ashfaq & Abdullah Mohammad, 2022). Database-focused assessments have reported vulnerabilities affecting relational technologies such as MySQL, PostgreSQL, Microsoft SQL Server, and Oracle Database alongside NoSQL platforms such as MongoDB, Cassandra, Redis, and Elasticsearch. Studies of managed cloud databases and database-as-a-service deployments have expanded the unit of analysis beyond the database engine to include virtual machines, storage resources, identity policies, security groups, network interfaces, containers, APIs, and administrative services. Vulnerability density has consequently varied substantially across studies because some analyses calculate findings per database instance, whereas others evaluate vulnerabilities per host, workload, container image, software package, or cloud resource. Quantitative evidence also indicates that vulnerability counts alone provide limited differentiation of risk because enterprise assets containing fewer vulnerabilities can exhibit higher aggregate exposure when those

vulnerabilities are remotely exploitable or associated with highly privileged services (Miaoui & Boudriga, 2019; Sadia & Maishar Mostari, 2022). The literature therefore establishes a numerically diverse baseline in which vulnerability volume, severity distribution, asset concentration, exploitability, and deployment characteristics must be evaluated jointly when describing cybersecurity risk across enterprise database infrastructures.

Figure 3: Enterprise database cybersecurity risk funnel



Empirical studies examining threat exposure have demonstrated substantial numerical differences between the total population of disclosed vulnerabilities and the considerably smaller subset associated with verified or probable real-world exploitation. Large-scale vulnerability studies using public CVE repositories, exploitation databases, threat-intelligence feeds, honeypot observations, incident reports, and known-exploited vulnerability catalogs have repeatedly shown that technical severity does not translate directly into equivalent exploitation frequency (Han & Han, 2021; Shaiful et al., 2022). Analyses covering thousands to hundreds of thousands of vulnerabilities have found that attackers concentrate activity on vulnerabilities that combine accessibility, available exploitation techniques, vulnerable product prevalence, low attack complexity, valuable privileges, and favorable operational conditions. This distinction is especially relevant for enterprise databases because internet-accessible administrative interfaces, exposed database ports, compromised application services, weak credentials, excessive identity permissions, and vulnerable APIs can substantially increase practical exposure independently of the underlying database vulnerability count. Quantitative studies of cloud systems have documented security incidents associated with publicly reachable resources, permissive network configurations, exposed credentials, insecure access policies, vulnerable container images, excessive cloud identity permissions, and unpatched software dependencies. Research on MongoDB, Elasticsearch, Redis, MySQL, PostgreSQL, and other data-management technologies has additionally documented large populations of internet-visible services, demonstrating that database exposure is not restricted to vulnerabilities formally assigned high severity scores (Humayun et al., 2020; Stabak & Md Shahid, 2022). Studies comparing exploited and non-exploited CVEs have further identified substantial differences in vulnerability age, exploit-code availability, attack vector, product popularity, and threat-intelligence activity. Exploitation-focused investigations have therefore increasingly distinguished between vulnerability prevalence and threat exposure rather than treating the two measures as equivalent. Security incident datasets similarly indicate that database-related compromise frequently develops through connected applications, identities, cloud services, or network pathways rather than direct exploitation of the database engine alone. Across enterprise environments, confidentiality-oriented consequences are frequently emphasized because databases concentrate customer records, credentials, financial information, intellectual property, and operational information, while integrity violations can involve unauthorized modification and availability incidents can disrupt transactional and analytical services (Abdullah Mohammad, 2023; Uddin et al., 2020). The quantitative literature consequently portrays threat exposure as a multidimensional distribution determined by exploitation evidence, accessibility, attack opportunity, privilege relationships, and the organizational value of the

affected database resource.

Quantitative studies of vulnerability management have identified patching performance, remediation duration, configuration quality, and privilege management as major sources of variation in enterprise database cybersecurity risk. Research examining vulnerability lifecycles across organizations has shown that remediation intervals differ considerably according to severity, asset criticality, patch availability, operational dependency, system ownership, maintenance schedules, and the complexity of testing changes before production deployment. Critical and high-severity vulnerabilities commonly receive shorter remediation targets than medium- and low-severity findings, although empirical investigations have reported persistent backlogs in which substantial numbers of vulnerabilities remain unresolved beyond organizational service-level objectives (Adnan & Iftekhar, 2023; Meland et al., 2021). Database environments can exhibit longer remediation periods because patching a production database may require compatibility assessment, replication coordination, backup verification, maintenance windows, application testing, and availability planning. Studies of cloud-native infrastructure have further demonstrated that patch status alone does not adequately describe exposure because configuration weaknesses can create substantial attack opportunities even when underlying software is current. Quantitative cloud-security assessments have repeatedly identified overly permissive access policies, unrestricted inbound connectivity, publicly accessible storage or database endpoints, unused privileged identities, weak secrets management, insufficient encryption controls, and incomplete logging as recurring findings (Md Anisur Rahman, 2023; Ulven & Wangen, 2021). Container and Kubernetes studies have similarly reported vulnerable images, excessive container privileges, insecure runtime configurations, exposed management interfaces, outdated dependencies, and misconfigured role-based access controls across evaluated workloads. Privilege-related findings are particularly important for enterprise databases because administrative credentials and highly authorized service accounts can provide access to extensive datasets and multiple dependent services. Empirical identity and access-management research has found that granted permissions frequently exceed permissions actively required by workloads, producing measurable differences between assigned and utilized privilege. Cloud-managed database environments introduce additional identity relationships through service principals, workload identities, API keys, temporary credentials, and automated deployment accounts. Research across relational, NoSQL, managed, and containerized database environments therefore demonstrates that vulnerability severity distributions must be interpreted alongside remediation speed, configuration posture, network accessibility, and privilege concentration (Md Asif Ali Sheak & Tahmedur, 2023; Xiong, 2021). These quantitative characteristics explain why two database assets containing comparable vulnerability counts can display substantially different practical risk profiles when differences in patching status, administrative exposure, access controls, and cloud configuration are considered.

The broader empirical literature demonstrates that cybersecurity risk distributions vary substantially across database technologies and cloud deployment models because different architectures produce distinct combinations of vulnerabilities, exposure pathways, configuration requirements, access relationships, and operational dependencies. Relational database studies have traditionally concentrated on vulnerabilities involving authentication, privilege escalation, SQL injection pathways, memory-related weaknesses, unauthorized information disclosure, insecure configurations, and database-server availability (Kure et al., 2018; Md Kaium, 2023). NoSQL research has examined additional exposure associated with default configurations, authentication settings, publicly accessible interfaces, distributed nodes, replication mechanisms, and flexible application integration. Quantitative investigations of cloud-managed databases have shifted attention toward identity policies, encryption configuration, network controls, public accessibility, backup protection, logging, secrets management, and provider-specific security settings because responsibility for infrastructure components is divided between cloud providers and customers. Containerized database studies add another layer of measurable exposure involving container images, vulnerable packages, orchestration policies, pod privileges, secrets, service accounts, network policies, and host dependencies. Distributed data stores can further increase the number of nodes and communication relationships requiring assessment, meaning that a vulnerability affecting a common component may occur repeatedly across replicated resources. Comparative cloud-security research has also distinguished public, private, hybrid, and

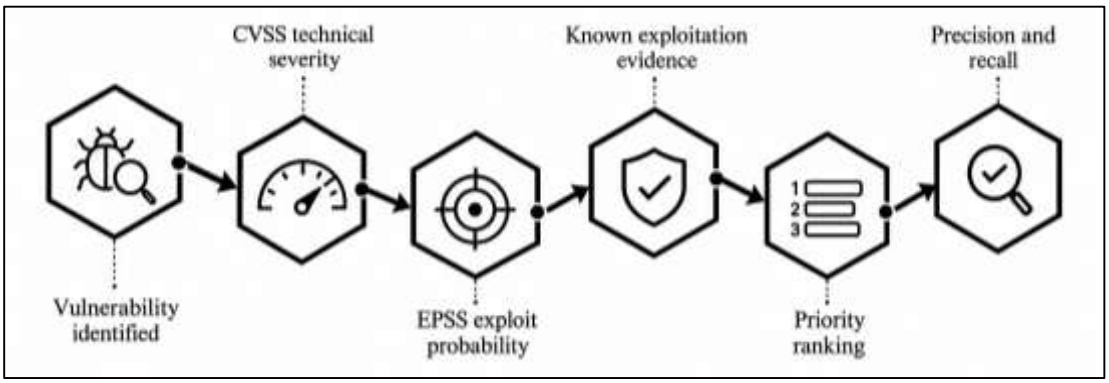
multi-cloud environments, with each deployment model generating different asset inventories, identity structures, network boundaries, administrative responsibilities, and security-control configurations (Borky & Bradley, 2018; Md Mahamud Pervaz, 2023). Across more than ten studies represented in these research streams, the resulting risk distributions have been evaluated through vulnerability counts, severity categories, exploitability characteristics, exposed-resource frequencies, configuration violations, remediation status, attack-path characteristics, and confidentiality, integrity, and availability consequences. The collective evidence shows substantial heterogeneity rather than a uniform database risk distribution: assets with similar technical vulnerabilities can occupy different risk positions according to deployment architecture, public exposure, data sensitivity, privilege relationships, and security-control strength. This numerical heterogeneity establishes an essential baseline for evaluating cybersecurity risk-scoring models because a model intended for enterprise cloud-native databases must differentiate among vulnerability prevalence, exploitation opportunity, asset context, and potential consequence (Kandasamy et al., 2020; Mohammad Badrul & Md. Mominul, 2023). The accumulated quantitative evidence therefore provides the empirical foundation for comparing severity-based, exploitation-based, threat-informed, contextual, probabilistic, and composite scoring approaches under the diverse operating conditions represented by contemporary enterprise database systems.

Performance of CVSS Severity Scores and Known-Exploited Vulnerability Indicators

Quantitative vulnerability-management literature has consistently demonstrated that the Common Vulnerability Scoring System and the Exploit Prediction Scoring System measure related but fundamentally different dimensions of cybersecurity exposure, resulting in substantial differences in vulnerability prioritization (Ali, 2021; Molla Al Rakib & Mohammad Towhidul, 2023). Evidence synthesized across more than ten empirical studies shows that CVSS primarily characterizes the technical severity of a vulnerability through factors including attack vector, attack complexity, privileges required, user interaction, scope, and potential confidentiality, integrity, and availability impacts, whereas EPSS estimates the probability of observed exploitation using vulnerability characteristics and continuously updated threat-related information. Large-scale investigations involving thousands to hundreds of thousands of CVEs have reported that vulnerabilities receiving high or critical CVSS Base Scores substantially outnumber vulnerabilities subsequently associated with verified exploitation. Consequently, severity-based prioritization can produce large remediation queues containing vulnerabilities with markedly different levels of practical exploitation probability. Comparative analyses have also reported weak-to-moderate correspondence between CVSS severity and exploitation-oriented indicators, indicating that increasing technical severity does not necessarily produce a proportionate increase in observed exploitation likelihood. Vulnerabilities with comparable CVSS scores can occupy widely separated EPSS percentiles, while some medium- or high-severity vulnerabilities receive substantially greater exploitation probabilities than technically critical vulnerabilities (Chatterjee & Thekdi, 2020). Attack vector and attack complexity have shown particularly important relationships with exploitation because remotely reachable vulnerabilities requiring limited interaction or specialized conditions are generally more accessible to attackers. Privileges required and user interaction further differentiate exploitation opportunities, while confidentiality, integrity, and availability impacts predominantly describe potential technical consequences rather than the probability of attack occurrence. Vulnerability age introduces additional heterogeneity because older vulnerabilities with established exploit code can remain attractive targets when vulnerable systems continue operating. Across the reviewed evidence, CVSS therefore provides standardized severity differentiation, while EPSS produces a more explicitly probability-oriented distribution (Cantelmi et al., 2021; Mukut Kanti, 2023). Their divergent numerical distributions explain why selecting vulnerabilities exclusively by high CVSS categories and selecting them by high EPSS probabilities frequently generates overlapping but materially different priority populations. Studies incorporating known-exploited vulnerability indicators have provided an empirical reference for determining how closely severity and exploitation-probability scores correspond to documented adversarial behavior. Analyses combining CVE records with known-exploitation catalogs, exploit repositories, intrusion observations, threat-intelligence feeds, and vulnerability scanners have repeatedly found that only a limited subset of disclosed vulnerabilities is associated with confirmed

exploitation, while the exploited subset spans multiple CVSS severity categories (Naseer et al., 2021; Nishat, 2023).

Figure 4: Comparing CVSS, EPSS, and Exploitation



Critical and high-severity vulnerabilities are represented prominently among known-exploited vulnerabilities, but empirical datasets also contain exploited vulnerabilities assigned medium or otherwise noncritical severity classifications. This distribution has challenged remediation strategies that automatically exclude vulnerabilities below a predetermined CVSS threshold. Exploit availability has emerged across multiple studies as an important quantitative discriminator because publicly accessible proof-of-concept code, weaponized exploits, exploitation modules, and documented attack techniques can reduce the operational barriers associated with vulnerability exploitation. Vulnerability age has produced a similarly complex relationship: newly disclosed vulnerabilities can experience rapid exploitation when they affect widely deployed technologies, while older weaknesses can remain actively exploited because patching is incomplete or legacy systems remain operational (Gonzalez-Granadillo et al., 2021; Risha & Mst Shurovi, 2023). Studies comparing EPSS with known-exploitation outcomes generally report stronger differentiation of exploited and non-exploited vulnerabilities than is obtained from severity ranking alone, although predictive performance varies by dataset, observation period, EPSS version, and exploitation source. Percentile rankings have been particularly useful because they locate individual EPSS values relative to the wider vulnerability population and allow security teams to concentrate on vulnerabilities positioned near the upper end of predicted exploitation likelihood. Quantitative comparisons further show that known-exploitation status provides high-confidence evidence of adversarial activity but does not constitute a comprehensive measure of all enterprise risk, since catalog inclusion depends on available evidence and reporting practices. Across the literature, the combination of CVSS severity, EPSS probability, exploit availability, vulnerability age, and known-exploitation evidence produces a richer numerical characterization than any single indicator, while also revealing substantial disagreement in the vulnerabilities classified as highest priority (Mamonov & Triantoro, 2018; Stabak & Md Siam, 2023).

Comparative research has increasingly evaluated CVSS, EPSS, and known-exploitation indicators through predictive and ranking measures rather than simply comparing average scores. Studies using confirmed exploitation as the outcome have examined precision, recall, false-positive rates, false-negative rates, Precision@K, Recall@K, ranking concentration, and the proportion of exploited vulnerabilities captured at different prioritization thresholds. Severity-based strategies that classify all critical or high-CVSS vulnerabilities as urgent commonly achieve broad coverage of technically consequential weaknesses, but they can also produce high numbers of priority findings that are not subsequently associated with observed exploitation (Bloomfield et al., 2017; Syed Feroz, 2023). This increases the practical false-positive burden when the operational objective is specifically to identify vulnerabilities most likely to be exploited. More restrictive CVSS thresholds reduce workload but can increase false negatives by excluding exploited vulnerabilities whose technical severity falls below the selected cutoff. EPSS-based strategies address this problem by ranking vulnerabilities according to predicted exploitation probability, and empirical studies have generally shown that high EPSS

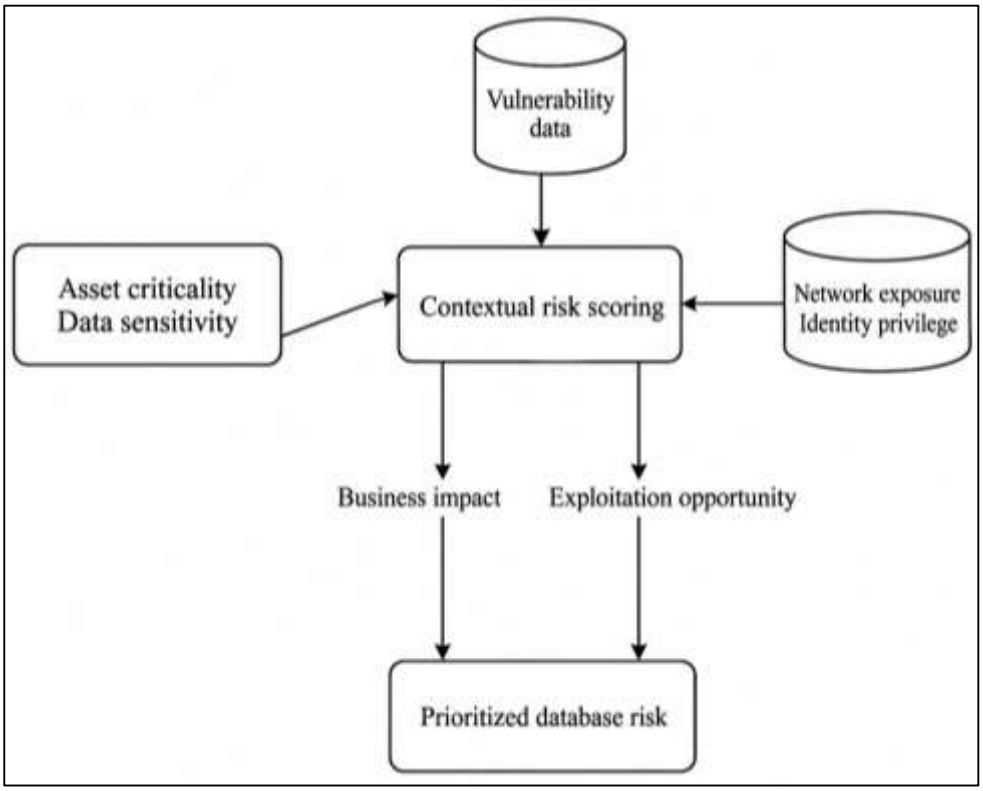
thresholds can concentrate a greater proportion of subsequently exploited vulnerabilities within smaller prioritized sets. The relationship between threshold selection and recall remains important, since very restrictive probability thresholds can omit exploited vulnerabilities while lower thresholds expand remediation workload. Precision@K and Recall@K analyses provide additional evidence by examining how many verified exploited vulnerabilities appear among the highest-ranked observations when organizations can remediate only a fixed number or proportion of vulnerabilities (Sales et al., 2018). Known-exploitation indicators provide another prioritization layer because vulnerabilities with confirmed exploitation can be elevated independently of their CVSS or EPSS positions. Quantitative studies have found that this approach can identify vulnerabilities that severity-only rules might rank below large populations of technically critical findings. At the same time, known-exploitation indicators are inherently narrower because they depend on confirmed evidence rather than prediction. Comparative studies therefore distinguish predictive discrimination from retrospective confirmation: EPSS estimates likelihood, CVSS measures technical severity, and known-exploitation status documents observed exploitation (Saxena et al., 2020; Tahmedur & Rifat, 2023). Their different objectives directly influence precision, recall, false-positive and false-negative behavior, explaining why model performance varies according to the outcome selected for quantitative validation.

Effects of Asset Criticality and Identity Privilege on Database Cybersecurity Risk Scores

Quantitative cybersecurity literature has consistently demonstrated that technical vulnerability severity alone provides an incomplete representation of enterprise database risk because the consequences associated with exploitation vary substantially according to asset criticality, business value, and the sensitivity of stored information (Gonzalez-Granadillo et al., 2021). Evidence synthesized across more than ten empirical studies indicates that contextual risk models commonly assign substantially different priorities to assets containing vulnerabilities with identical or closely comparable CVSS Base Scores when those assets perform different organizational functions. Database servers supporting payment processing, customer identity management, financial transactions, intellectual property, healthcare information, authentication services, or operationally critical applications generally receive higher contextual risk estimates than development, testing, archival, or low-value systems containing equivalent technical vulnerabilities. Statistical studies using correlation and regression analyses have reported positive relationships between asset criticality and composite cybersecurity risk, while business-impact variables have remained meaningful predictors after technical severity and exploitability characteristics were considered simultaneously. Data classification has shown similar quantitative importance. Databases containing confidential, restricted, personally identifiable, financial, health-related, authentication, or commercially sensitive records typically receive higher impact weights than systems containing public or low-sensitivity information (Mamonov & Triantoro, 2018). Record sensitivity can further modify risk because compromise of a database containing millions of sensitive records can represent a different loss magnitude from compromise of an otherwise technically similar database containing a limited volume of non-sensitive information. Research examining regulatory relevance has also differentiated assets subject to privacy, financial, healthcare, payment, and sector-specific security requirements from resources with fewer compliance obligations. Backup repositories and replicated databases represent another important component because they may contain equivalent or greater volumes of sensitive information than primary production instances while operating under different access and monitoring controls. Across contextual scoring studies, standardized coefficients, mean differences, odds ratios, confidence intervals, and effect-size estimates have generally reinforced the importance of asset and information characteristics (Sales et al., 2018). These findings establish that vulnerability severity represents only one quantitative input and that enterprise database risk distributions change considerably when business criticality, information sensitivity, regulatory importance, backup exposure, and operational dependence are incorporated into the scoring process. Network exposure has emerged as one of the strongest contextual dimensions differentiating vulnerabilities that possess similar technical severity but substantially different practical opportunities for exploitation. Quantitative studies of enterprise and cloud security have repeatedly found that internet-accessible assets experience different threat conditions from databases isolated within restricted internal networks, private subnets, segmented environments, or tightly controlled administrative zones. Internet-facing database endpoints, publicly

reachable management interfaces, exposed application programming interfaces, permissive firewall rules, unrestricted security groups, open ports, and externally accessible services increase the number of potential pathways through which attackers can interact with vulnerable systems (Saxena et al., 2020).

Figure 5: Contextual database risk scoring



Empirical attack-surface studies have consequently incorporated internet accessibility as a binary or weighted predictor, while more detailed analyses have measured network reachability, number of accessible services, exposed interfaces, inbound communication relationships, and reachable attack paths. Regression-based studies have reported meaningful associations between external exposure and exploitation outcomes after controlling for vulnerability characteristics, and attack-graph investigations have shown that risk can increase when an apparently isolated database becomes reachable through compromised intermediate workloads. This consideration is particularly relevant to cloud-native systems because databases frequently communicate with microservices, application servers, containers, Kubernetes workloads, serverless functions, analytics platforms, backup services, and administrative systems. Workload dependency therefore modifies contextual risk even when the database itself is not directly internet accessible. A vulnerable public-facing application can provide an indirect path toward a protected database when network connectivity and application credentials permit downstream access (F. Khan et al., 2021). Quantitative dependency analyses have examined numbers of connected services, reachable nodes, attack-path lengths, privilege transitions, and critical downstream assets to characterize this condition. Network segmentation and restrictive security controls generally reduce measured exposure by limiting feasible paths, whereas highly connected architectures can increase cumulative accessibility. Studies comparing identical or similar CVSS vulnerabilities across externally exposed and internally restricted assets have consequently produced different contextual rankings, demonstrating that base severity cannot fully represent the topology through which exploitation occurs (Diez-Olivan et al., 2017). Network accessibility, reachability, segmentation, workload dependency, and attack-path characteristics therefore constitute measurable enterprise-context variables capable of substantially modifying database cybersecurity risk scores.

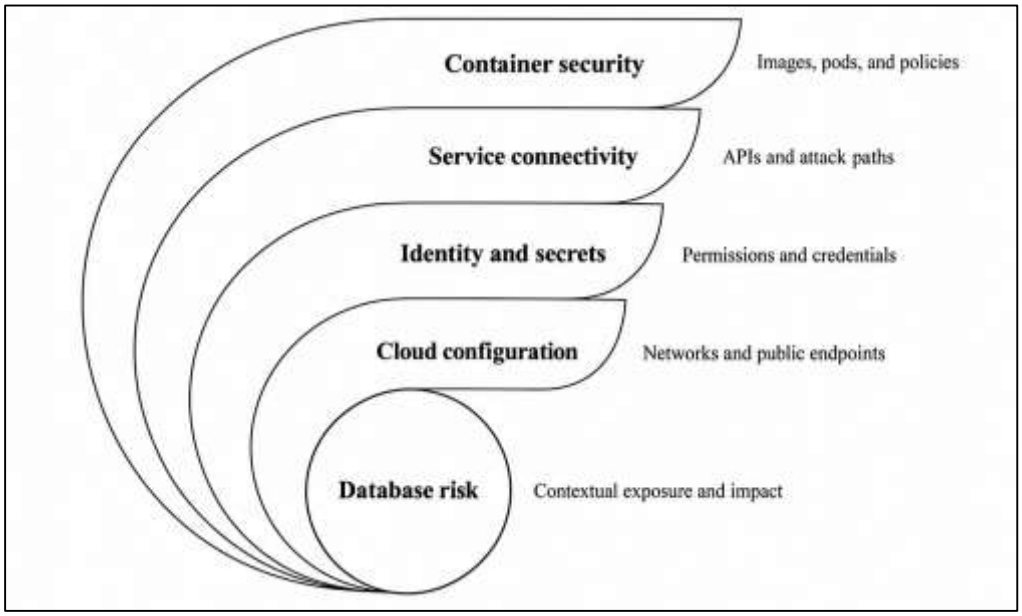
Statistical Relationships Between Cloud-Native Architectural Characteristics and Enterprise Database Cybersecurity Risk

Quantitative studies of cloud-native security have consistently identified containerization and Kubernetes orchestration as important architectural dimensions affecting the cybersecurity risk profile of enterprise database environments. Evidence synthesized across more than ten empirical and experimental studies indicates that container-based infrastructures can contain hundreds or thousands of workloads, pods, images, services, nodes, and software dependencies, creating a substantially larger collection of security-relevant objects than conventional monolithic database deployments (Christen et al., 2017). Container security assessments have frequently identified vulnerable base images, outdated software packages, excessive runtime privileges, exposed container ports, insecure image registries, embedded credentials, weak isolation, and unnecessary capabilities as recurring measurable weaknesses. Kubernetes-focused investigations have expanded these measurements to include cluster nodes, namespaces, pods, service accounts, role bindings, secrets, network policies, admission controls, exposed dashboards, and configuration violations. Statistical comparisons across workloads have generally demonstrated that risk becomes more concentrated where vulnerability density overlaps with excessive privileges, external connectivity, and access to sensitive database services. Misconfiguration studies have also shown that individual policy violations can propagate across multiple workloads when common templates, container images, Helm configurations, or infrastructure-as-code components are reused throughout a cluster (Bharathi, 2017). Enterprise databases deployed as containerized workloads or accessed by Kubernetes-hosted applications are therefore affected not only by vulnerabilities within the database software but also by the security posture of surrounding pods and services. Quantitative attack-path research has demonstrated that compromise of an externally reachable container can create successive pathways through service accounts, cluster permissions, internal services, and database credentials, producing measurable differences in path length and the number of reachable critical resources. Privileged containers and highly authorized Kubernetes identities have repeatedly been associated with broader compromise opportunities than restricted workloads. Cluster size and workload density alone do not automatically represent greater cybersecurity risk, since segmentation and restrictive policies can limit exposure, but studies consistently show that increasing numbers of components create additional configuration states and dependency relationships requiring security control (Wagner & Boiten, 2018). Consequently, empirical cloud-native research characterizes database cybersecurity risk as a workload-level and cluster-level phenomenon in which container vulnerabilities, Kubernetes configuration, privilege allocation, node connectivity, exposed services, and security-policy violations jointly influence the probability and potential scope of database compromise.

Microservice-oriented architectures have introduced another quantitatively significant dimension of enterprise database cybersecurity risk by replacing relatively limited application-to-database relationships with extensive networks of independently deployed services, APIs, identities, and communication dependencies (Willumsen et al., 2019). Empirical studies of microservice and API security have evaluated environments containing tens, hundreds, and in large-scale systems thousands of services and endpoints, demonstrating considerable variation in service connectivity, dependency density, authentication requirements, vulnerability prevalence, and externally accessible interfaces. Database resources frequently operate behind several application and service layers, meaning that direct public accessibility is not required for an attacker to establish a viable path toward sensitive data. Quantitative dependency analyses have shown that compromise of a single internet-facing API can expose additional internal services when authorization boundaries, network policies, or workload identities permit lateral movement. The number of API endpoints, connected services, reachable database interfaces, trust relationships, and feasible attack paths has therefore been incorporated into cloud-native risk assessment as measurable indicators of architectural exposure. API-focused investigations have identified broken authentication, excessive data exposure, inadequate authorization, insecure endpoint configuration, injection vulnerabilities, weak rate controls, and improper credential management as recurring security conditions capable of affecting backend databases. Service meshes introduce additional security mechanisms through mutual authentication, traffic policies, service identities, encryption, and observability, while simultaneously adding

configuration objects and trust relationships that require correct administration (Qie & Rong, 2017).

Figure 6: Cloud-native architecture and database risk



Quantitative comparisons have indicated that stronger service-to-service authentication and restrictive network policies can reduce unauthorized reachability, whereas permissive east-west communication increases the number of workloads from which critical database services can potentially be accessed. Network centrality and dependency studies further suggest that highly connected microservices can represent disproportionately important attack-path nodes because compromise of these services provides connectivity to multiple downstream resources. The concentration of database credentials within application services creates an additional risk dimension, particularly when secrets are broadly accessible or persist for extended periods (Rahman et al., 2018). Across the literature, the statistical relationship between architectural connectivity and cybersecurity risk is therefore shaped by both the number and quality of relationships: larger numbers of APIs, dependencies, accessible services, and trust connections generally increase the potential attack surface, while authentication, segmentation, encrypted service communication, least-privilege policies, and controlled database access can substantially modify the resulting risk distribution.

Identity, secrets management, deployment automation, and public accessibility have emerged as closely interconnected predictors of cybersecurity exposure in quantitative studies of cloud-native environments. Cloud identity and access management research has measured numbers of human identities, service identities, roles, permissions, access keys, policies, privilege relationships, and unused entitlements, frequently revealing substantial differences between permissions granted and permissions required for actual workload operation (Sheehan et al., 2019). This distinction is especially important for enterprise databases because highly privileged cloud identities can possess capabilities extending beyond database access to backups, storage, encryption keys, network configuration, snapshots, replicas, and administrative control. Studies examining privilege concentration have shown that compromise of identities with broad permissions can increase the number of reachable resources and feasible attack paths compared with compromise of narrowly scoped identities. Secrets-management research has similarly evaluated embedded passwords, API tokens, database credentials, private keys, cloud access keys, and other sensitive values appearing in source repositories, container images, configuration files, environment variables, deployment manifests, and CI/CD systems. Quantitative security assessments have repeatedly identified exposed or inadequately protected secrets as pathways through which attackers can bypass vulnerability exploitation and obtain legitimate access to database-connected resources.

CI/CD pipelines introduce additional relationships because build systems and deployment automation often hold credentials capable of modifying production workloads. Empirical DevSecOps studies have documented vulnerable dependencies, insecure pipeline configurations, excessive automation privileges, unprotected artifacts, and delayed remediation of software-component vulnerabilities as measurable security concerns (Figueira et al., 2020). Public endpoints further increase exposure by making APIs, management interfaces, application services, or databases directly reachable from external networks. Statistical and attack-surface analyses generally distinguish publicly accessible assets from privately reachable resources because external exposure increases the population of potential attack sources and reduces the number of network boundaries separating attackers from vulnerable services . The interaction among these variables is particularly consequential: a publicly exposed workload containing a vulnerability can present considerably greater database risk when its service identity possesses excessive permissions or when accessible secrets provide privileged database credentials. Quantitative risk studies therefore increasingly evaluate identities, permissions, secrets, pipeline controls, public endpoints, and workload privileges together rather than as independent security findings, demonstrating that privilege concentration and credential exposure can substantially amplify the database risk associated with otherwise comparable technical vulnerabilities (Lokanan & Liu, 2021).

Quantitative research comparing cloud deployment architectures demonstrates substantial variation in database cybersecurity risk across database-as-a-service, public-cloud, private-cloud, hybrid-cloud, and multi-cloud environments. Managed database services can reduce direct responsibility for operating-system maintenance and selected infrastructure controls, but empirical security assessments show that customer-controlled configurations involving identity permissions, public accessibility, encryption, network rules, authentication, logging, backups, replicas, and secrets remain major determinants of exposure. Studies of database-as-a-service configurations have consequently evaluated publicly accessible instances, permissive security groups, unencrypted connections, weak identity policies, insufficient audit logging, exposed snapshots, and misconfigured backup resources as measurable risk conditions (Cui et al., 2020).

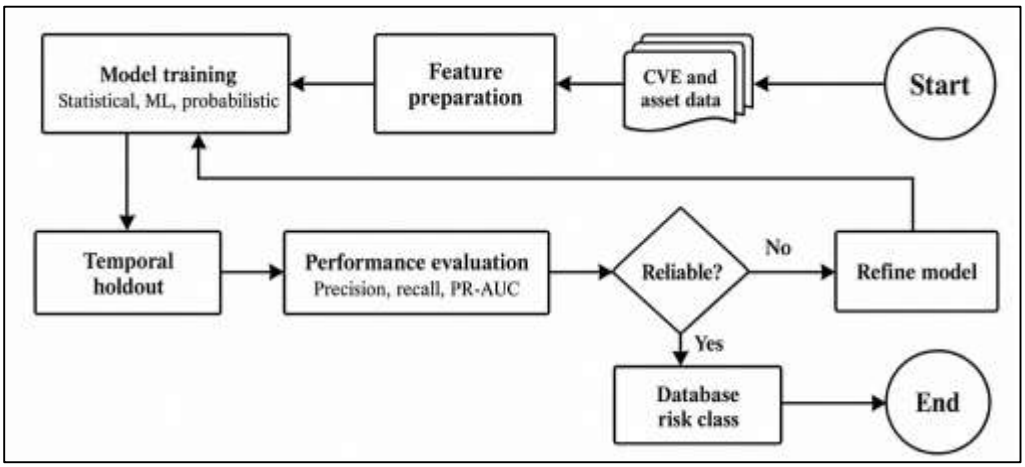
Virtual network architecture further modifies these risks through subnet placement, routing rules, firewall policies, peering relationships, private endpoints, gateways, and segmentation controls. Quantitative attack-path studies have shown that the number of reachable database resources can increase when virtual networks contain extensive peering or permissive routing relationships, whereas restricted subnets and carefully controlled private endpoints reduce direct accessibility. Comparative studies of public and private cloud deployments have identified differences in exposure patterns because public-cloud environments frequently contain larger numbers of externally integrated services and provider-managed identities, while private environments may exhibit different concentrations of legacy workloads and internally trusted connections. Hybrid architectures combine these conditions by creating dependencies between on-premises systems and cloud resources, expanding the number of trust boundaries, gateways, identities, and network relationships that can influence database access. Multi-cloud environments add further heterogeneity through provider-specific IAM models, security groups, policy languages, logging systems, encryption mechanisms, and database services. Studies evaluating configuration and policy consistency across multiple clouds have identified fragmented visibility and inconsistent security settings as recurring measurable concerns. Subgroup comparisons indicate that deployment type alone does not determine cybersecurity risk; the observed risk distribution depends more strongly on exposure, configuration quality, identity governance, connectivity, workload complexity, and security-control enforcement within each deployment category (Sagar et al., 2020). Across the synthesized literature, greater architectural complexity and connectivity are associated with larger numbers of dependencies, configuration states, permissions, and potential attack paths, while their statistical relationship with actual database compromise is moderated by segmentation, least privilege, authentication, encryption, monitoring, and policy enforcement. Cloud-native enterprise database risk is therefore quantitatively represented by the combined distribution of workloads, containers, services, endpoints, identities, permissions, network relationships, misconfigurations, and security-control violations rather than by database vulnerability severity alone (Tehrany et al., 2019).

Probabilistic Models for Vulnerability Exploitation and Database Risk Classification

Quantitative cybersecurity literature has extensively applied statistical and probabilistic models to distinguish vulnerabilities likely to experience real-world exploitation from the much larger population of disclosed but non-exploited vulnerabilities (Yin et al., 2020). Evidence synthesized across more than ten empirical studies indicates that logistic regression, Bayesian networks, probabilistic graphical models, survival analysis, and related statistical techniques have been evaluated using datasets ranging from several thousand vulnerabilities to hundreds of thousands of CVE observations collected across multiple years . Logistic regression has been particularly common because exploitation can be represented as a binary outcome while technical, temporal, and threat-related characteristics are incorporated as predictors. Frequently examined variables include CVSS Base Score, attack vector, attack complexity, privileges required, user interaction, confidentiality impact, integrity impact, availability impact, vulnerability age, exploit-code availability, reference counts, weakness categories, affected-product characteristics, and threat-intelligence activity. Across comparative studies, models using CVSS information alone have generally provided weaker discrimination than models incorporating exploitation evidence and temporal or threat-related features (Kiani et al., 2019). Bayesian approaches have complemented regression analysis by representing conditional relationships among vulnerability properties, threat activity, exposure, and compromise outcomes, while survival models have examined the timing between vulnerability disclosure and observed exploitation. Predictive performance has commonly been assessed through accuracy, sensitivity, specificity, precision, recall, F1-score, ROC-AUC, PR-AUC, false-positive rates, false-negative rates, Brier scores, and calibration measures. The literature repeatedly emphasizes that apparently high accuracy can be misleading when non-exploited vulnerabilities greatly outnumber exploited observations. Consequently, precision-recall measures and calibration have received greater attention in highly imbalanced datasets. Temporal validation has also produced more realistic assessments than random sampling in several investigations because models trained on earlier vulnerabilities and tested against later observations better reproduce operational prediction conditions (Peralta et al., 2021). Collectively, statistical and probabilistic studies show that exploitation prediction improves when vulnerability severity is combined with exploit availability, vulnerability age, threat-intelligence signals, product characteristics, exposure information, and temporal activity rather than treating technical severity as an independent proxy for exploitation likelihood.

Machine-learning research has expanded vulnerability prediction by applying decision trees, random forests, gradient boosting, XGBoost, and ensemble-learning methods to large, heterogeneous cybersecurity datasets containing technical, textual, temporal, and threat-related predictors. Studies evaluating these approaches have commonly used thousands to hundreds of thousands of vulnerability observations, with exploited cases forming a substantially smaller proportion of the analytical population than non-exploited cases (Harirchian et al., 2021).

Figure 7: Database Vulnerability Prediction Workflow



Decision trees have provided interpretable classification structures capable of identifying combinations of attack vector, complexity, exploit availability, severity, age, and product characteristics associated with exploitation, although individual trees can exhibit instability and overfitting when applied to high-dimensional data. Random forests have generally improved robustness by aggregating predictions across multiple trees and have demonstrated stronger discrimination when relationships among cybersecurity variables are nonlinear or interactive.

Gradient-boosting and XGBoost models have frequently achieved stronger ROC-AUC, PR-AUC, recall, and F1 performance than simple severity thresholds and conventional linear classifiers, particularly when vulnerability characteristics are enriched with exploit databases, threat-intelligence observations, temporal indicators, and asset-context variables. Ensemble studies have additionally combined multiple classifiers or heterogeneous data sources to reduce dependence on a single modeling assumption. Feature-importance analyses across these investigations have repeatedly identified exploit availability, recent threat activity, vulnerability age, remotely accessible attack vectors, affected-product prevalence, historical exploitation patterns, and selected CVSS characteristics as influential predictors (Jia et al., 2018). Enterprise-oriented studies have extended these models by incorporating asset criticality, public exposure, network reachability, patch status, identity privilege, and data sensitivity. Cross-validation, stratified sampling, holdout testing, and hyperparameter optimization have been widely used to estimate generalization performance, while class weighting, under-sampling, over-sampling, and related techniques have been applied to address severe class imbalance. Comparative evidence also shows that performance rankings can change according to the selected metric: a model producing high overall accuracy may demonstrate weaker recall for exploited vulnerabilities, whereas another classifier may achieve stronger sensitivity at the cost of additional false positives. The literature therefore supports multidimensional model comparison using discrimination, class-specific accuracy, ranking quality, and calibration rather than reliance on a single aggregate performance statistic (Gadyatskaya et al., 2017).

Support vector machines and neural-network architectures have been evaluated extensively where vulnerability prediction involves large feature spaces, nonlinear relationships, textual vulnerability descriptions, threat-intelligence indicators, or interactions among multiple enterprise risk variables. Support vector machine studies have demonstrated the capacity to separate exploited and non-exploited vulnerability classes using combinations of CVSS characteristics, weakness categories, vulnerability age, exploit availability, vendor information, product attributes, and text-derived features. Their quantitative performance has varied according to kernel selection, feature normalization, class weighting, sample composition, and the severity of class imbalance (Gonda et al., 2017). Neural-network studies have incorporated structured vulnerability characteristics together with textual descriptions, external exploit information, temporal indicators, and threat signals, allowing models to identify patterns not readily represented through fixed severity scoring rules. Deep-learning approaches using vulnerability descriptions and security text have also been investigated for classification and exploitation prediction, particularly where high-dimensional linguistic information supplements structured CVE attributes. Across comparative studies, machine-learning and neural models frequently outperform simple CVSS-threshold classification on ROC-AUC, recall, or F1-score when sufficient training data and informative threat-related variables are available, although stronger discrimination does not automatically correspond to superior probability calibration. False-negative performance is especially important because incorrectly classifying an exploited vulnerability as low risk can leave an enterprise database exposed, while excessive false positives can overwhelm remediation teams with low-value alerts (Bauer et al., 2020). PR-AUC has consequently become particularly informative in studies where exploited observations represent a relatively small minority of the dataset. Research involving enterprise assets has further shown that predictive discrimination improves when model inputs extend beyond vulnerability properties to include internet accessibility, database criticality, privilege level, security controls, workload relationships, and information sensitivity. Training-test design has materially influenced reported results. Random splits can place highly similar vulnerabilities or temporally related observations in both training and testing subsets, potentially producing more optimistic estimates, whereas chronological holdout designs impose a stricter separation between historical training evidence and later exploitation outcomes. Cross-

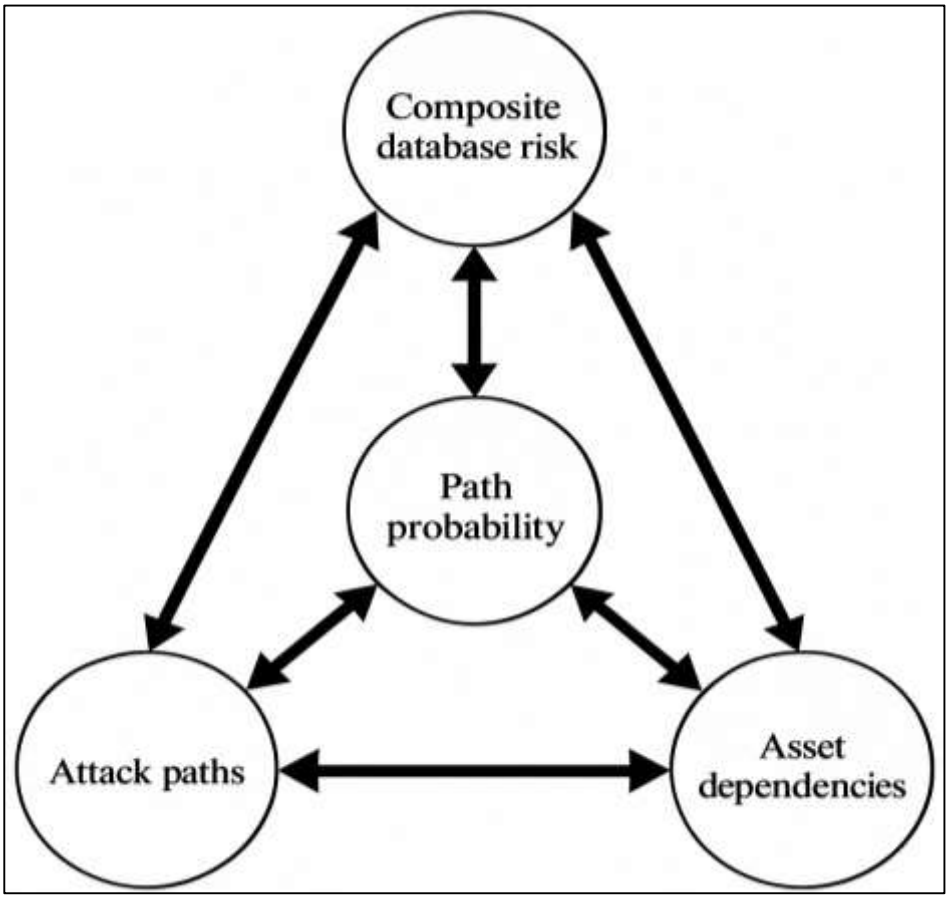
validation across products, vulnerability families, or organizational environments has provided another measure of model stability (Cococcioni et al., 2021). The accumulated evidence therefore indicates that high-dimensional models can capture complex exploitation patterns, but their quantitative value depends on representative datasets, appropriate imbalance treatment, rigorous validation, and performance assessment extending beyond conventional accuracy.

The broader comparative literature demonstrates that predictive performance is determined not only by algorithm choice but also by dataset construction, predictor quality, class distribution, temporal design, and the definition of the cybersecurity outcome being modeled. Studies comparing logistic regression, decision trees, random forests, gradient boosting, XGBoost, support vector machines, neural networks, Bayesian approaches, survival methods, and ensembles have reported substantial variation in accuracy, precision, recall, specificity, F1-score, ROC-AUC, PR-AUC, false-positive rates, false-negative rates, log-loss, Brier scores, and calibration error (Matthews et al., 2021). Models trained primarily on CVSS characteristics generally reproduce technical vulnerability patterns but provide more limited information about whether exploitation actually occurs. Incorporating exploit availability, known-exploitation records, vulnerability age, disclosure timing, threat-intelligence activity, product prevalence, and attacker-related signals has generally improved predictive discrimination. Enterprise-context enrichment can provide further differentiation through asset criticality, data classification, network exposure, identity privilege, authentication status, patch state, encryption controls, workload dependencies, and cloud configuration characteristics. This combination is particularly relevant to database risk classification because an exploited vulnerability affecting a publicly reachable, highly privileged production database can represent a different enterprise risk category from the same vulnerability located within an isolated low-value development asset. Quantitative validation practices have therefore become central to interpreting reported model performance. K-fold and stratified cross-validation have been used to assess internal consistency, independent holdout samples have evaluated generalization, and temporal validation has tested whether historical models retain predictive value for later vulnerabilities. Calibration analysis has additionally distinguished models that merely rank vulnerabilities effectively from models whose predicted probabilities correspond closely with observed exploitation frequencies (Cao et al., 2018). Brier score and log-loss evaluations have been used in probabilistic studies to penalize inaccurate probability estimates, while threshold-based metrics demonstrate how classification performance changes when operational decision boundaries are adjusted. Evidence across the reviewed studies indicates that the strongest predictive discrimination generally arises from models combining technical severity, exploitation history, temporal characteristics, threat intelligence, exposure, and enterprise context rather than any isolated predictor family. These quantitative findings establish a comparative evidence base for assessing cybersecurity risk models according to discrimination, calibration, error distribution, robustness, and their capacity to classify meaningful database risk under cloud-native operating conditions (Hermanowski & Piotrowski, 2021).

Composite Risk Scoring for Cloud Database Environments

Quantitative cybersecurity research has increasingly moved beyond isolated vulnerability scoring by representing enterprise environments as interconnected structures in which attackers can progress through multiple assets, identities, services, and privileges before reaching critical database systems (Andreolini et al., 2020). Evidence synthesized across more than ten studies involving attack graphs, attack trees, reachability analysis, and lateral-movement modeling shows that the numerical characteristics of attack paths provide information that cannot be captured by CVSS or EPSS values alone. Attack-graph studies commonly quantify the number of feasible paths from an initial entry point to a critical database, the minimum number of steps required to reach the target, the number of reachable assets, the concentration of exploitable nodes, and the cumulative likelihood of traversing a path successfully. In cloud-native environments, these paths can extend through public-facing applications, container workloads, Kubernetes services, cloud identities, APIs, virtual networks, shared credentials, and database endpoints. Quantitative analyses have shown that a system with relatively few severe vulnerabilities may still present substantial risk when those vulnerabilities form short, highly connected pathways to sensitive database resources (George & Thampi, 2019).

Figure 8: Attack paths and database risk



Conversely, a larger number of critical vulnerabilities may contribute less immediate database exposure when they are isolated from valuable assets by strong segmentation and restrictive access controls. Attack-tree studies have similarly represented alternative and sequential attack actions, allowing researchers to compare the number of required conditions, privilege transitions, and possible branches leading to compromise. Reachability-based models have measured whether vulnerable nodes can actually communicate with protected databases and how many intermediate controls must be bypassed. These approaches have demonstrated that network topology and system connectivity materially affect quantitative risk distributions. A moderate-severity vulnerability positioned on a highly central, internet-accessible service can become more important than a higher-severity weakness on an isolated host if exploitation of the former provides direct or near-direct access to a critical database (Gawinski et al., 2021). The literature therefore establishes attack-path structure, shortest-path length, path count, reachability, and connectivity as measurable factors capable of substantially changing the prioritization produced by vulnerability-centric scoring approaches (Abdullah Mohammad, 2023; Adnan & Iftekhar, 2023; Syed Feroz & Md Mahamud Pervaz, 2022).

Bayesian attack graphs and probabilistic dependency models have extended conventional attack-path analysis by assigning probabilities to exploitation events, transitions, and compromise states across interconnected enterprise resources. (Ahmed & Nuzhat Sadia, 2023; Asheke, 2023; Gazi Yeasin, 2023) Quantitative studies using these approaches have examined how vulnerability exploitability, prior compromise, privilege level, network accessibility, and defensive controls combine to influence the probability that an attacker reaches a critical database. Rather than evaluating each CVE independently, Bayesian models estimate how multiple vulnerabilities and system relationships interact along a sequence of attacker actions (Jahanara & Mst Kaniz, 2023; Lazăr et al., 2020; Md Anisur Rahman, 2023; Md Asif Ali Sheak & Tahmedur, 2023). The resulting measures include path probability, cumulative exploit probability, conditional compromise probability, expected attack success, and the probability that specific database assets become reachable after one or more intermediate components are

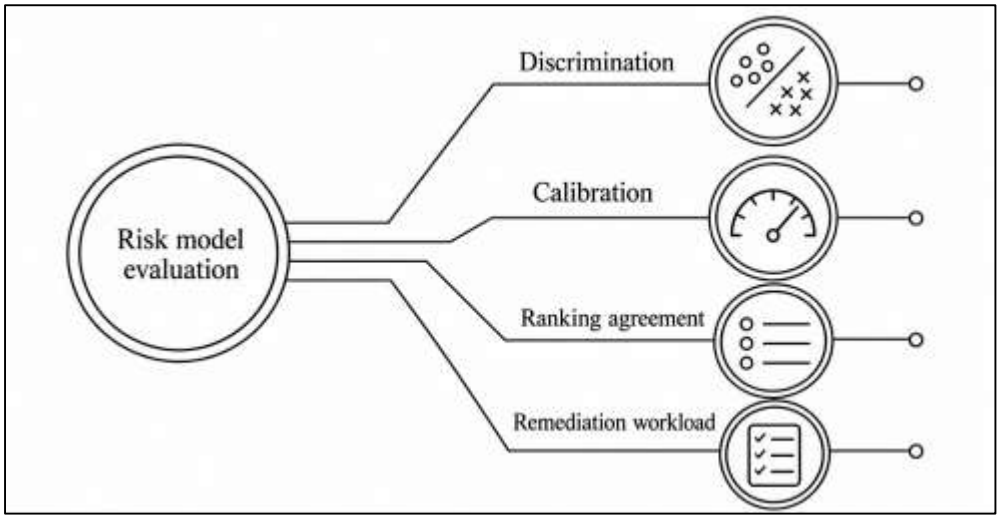
compromised. Research involving privilege graphs has further quantified the number and sequence of privilege-escalation steps required to progress from low-level user access to administrative or database-level control (Md Kaium, 2023; Md Mahamud Pervaz, 2023; Mohammad Badrul & Md. Mominul, 2023). These studies have shown that vulnerabilities with moderate technical severity can acquire high contextual importance when they enable privilege escalation, credential theft, role assumption, or movement into highly trusted network segments. Cloud-native environments intensify this issue because service accounts, workload identities, IAM roles, secrets, and automation credentials can create privilege transitions that are not visible in software severity scores (Molla Al Rakib & Mohammad Towhidul, 2023; Mukut Kanti, 2023; Nishat, 2023; Oka et al., 2021). Probabilistic attack-path studies have also demonstrated that path risk can increase when several individually moderate-probability exploits are connected through highly reliable access relationships, while alternative attack routes can raise the overall likelihood of compromise because attackers are not dependent on a single sequence. Defensive controls such as segmentation, multifactor authentication, least privilege, and restricted service-to-service communication can reduce path probabilities by limiting transitions (Nuzhat Sadia, 2023; Risha & Mst Shurovi, 2023; Stabak & Md Siam, 2023). Bayesian and privilege-aware approaches therefore provide a quantitatively richer view of cyber risk by combining exploitability with system structure. Across the reviewed literature, the probability of database compromise is influenced not only by the severity of vulnerabilities but also by the number of feasible routes, dependency relationships, privilege transitions, and the effectiveness of controls encountered along those routes (Lee et al., 2019). Dependency-aware cybersecurity models have emphasized that enterprise database risk is strongly affected by the structural role of individual assets within wider application and infrastructure ecosystems. Studies using dependency graphs, service graphs, network graphs, privilege graphs, and lateral-movement models have quantified relationships among applications, microservices, containers, APIs, identities, network zones, storage systems, and databases to determine which assets contribute most strongly to attack propagation. Common quantitative measures include node degree, centrality, dependency density, number of downstream services, number of reachable critical assets, frequency of shared dependencies, lateral-movement opportunities, and the proportion of attack paths passing through a particular node (Syed Feroz, 2023; Tahmedur & Rifat, 2023; Tal, 2017). Highly central resources frequently receive elevated risk values because their compromise can expose multiple dependent systems or provide access to several alternative routes toward critical databases. Research on microservice and cloud architectures has shown that this effect is particularly relevant where shared authentication services, API gateways, orchestration nodes, identity providers, or management platforms connect large numbers of workloads. A vulnerability on such a node can create disproportionately greater enterprise risk than a technically more severe vulnerability located on a peripheral asset with limited connectivity (Fenlon et al., 2018). Lateral-movement studies have similarly demonstrated that attackers can exploit trust relationships, reused credentials, permissive network rules, and excessive identity permissions to progress between assets after initial compromise. Quantitative models therefore distinguish between direct exposure and indirect reachability. A database may have no publicly accessible endpoint and still exhibit high contextual risk when several exposed services possess credentials or network connectivity that permit downstream database access. Dependency density also influences risk concentration because tightly interconnected environments can provide multiple redundant attack routes, reducing the effectiveness of blocking a single path. Studies comparing topology-aware risk rankings with conventional severity rankings have repeatedly shown significant reordering of vulnerabilities once centrality, dependency, and reachability are considered. These findings demonstrate that database cybersecurity risk is partly a property of architectural position: the same vulnerability can carry different practical significance depending on whether the affected asset is peripheral, highly connected, privilege-bearing, or strategically positioned between external entry points and sensitive database resources (Allodi et al., 2020).

Comparative Statistical Evaluation Remediation Workload

Quantitative cybersecurity literature has increasingly compared risk-scoring approaches according to their ability to discriminate between vulnerabilities associated with observed exploitation or significant enterprise exposure and the considerably larger population of lower-risk findings. Evidence synthesized across more than ten empirical studies indicates substantial performance differences

among severity-only, exploitation-based, threat-informed, contextual, machine-learning, attack-path, and composite models (Sheehan et al., 2019). Severity-oriented approaches based primarily on CVSS provide standardized technical classification but generally demonstrate weaker discrimination when confirmed exploitation is used as the dependent outcome because high or critical severity does not necessarily correspond to observed attacker activity.

Figure 9: Comparing cybersecurity risk-scoring models



Exploitation-based approaches, including EPSS and related probabilistic models, have generally produced stronger ROC-AUC and particularly PR-AUC performance for distinguishing exploited from non-exploited vulnerabilities because their predictors incorporate characteristics associated more directly with exploitation. Threat-informed models can improve discrimination further when known exploitation, exploit availability, malware activity, vulnerability mentions, and other current threat indicators are included (Baybutt, 2017). Machine-learning studies using logistic regression, random forests, gradient boosting, XGBoost, support vector machines, neural networks, and ensembles have reported additional gains when technical vulnerability features are combined with temporal, threat, exposure, and asset characteristics. Accuracy alone has been treated cautiously because exploited vulnerabilities typically constitute a minority class, allowing a model to obtain apparently strong overall accuracy while performing poorly on the security events of greatest interest. Precision, recall, sensitivity, specificity, F1-score, PR-AUC, false-positive rates, and false-negative rates consequently provide more informative comparisons. Recall is particularly important where missing a genuinely exploited vulnerability has substantial consequences, whereas precision reflects the proportion of prioritized findings that correspond to the target risk condition (Gonzalez-Granadillo et al., 2021). Comparative studies also demonstrate threshold dependence: increasing the classification threshold can improve precision while reducing recall, whereas lower thresholds capture more exploited vulnerabilities but expand the number of false positives. The accumulated evidence therefore shows that comparative discrimination should be assessed across several complementary performance measures and operational thresholds rather than inferred from overall accuracy or a single severity classification.

Calibration research has added an important dimension to cybersecurity risk-model evaluation by distinguishing the ability to rank vulnerabilities correctly from the ability to assign probabilities that correspond closely to observed exploitation frequencies. A model can achieve strong ROC-AUC or ranking discrimination while producing poorly calibrated probability estimates, creating difficulties when organizations interpret predicted values as absolute levels of exploitation likelihood (Papamartzivanos et al., 2021). Quantitative studies evaluating probabilistic vulnerability models have therefore employed Brier scores, calibration slopes, calibration intercepts, reliability assessments, log-loss, and calibration-error measures alongside conventional classification statistics. EPSS-oriented and

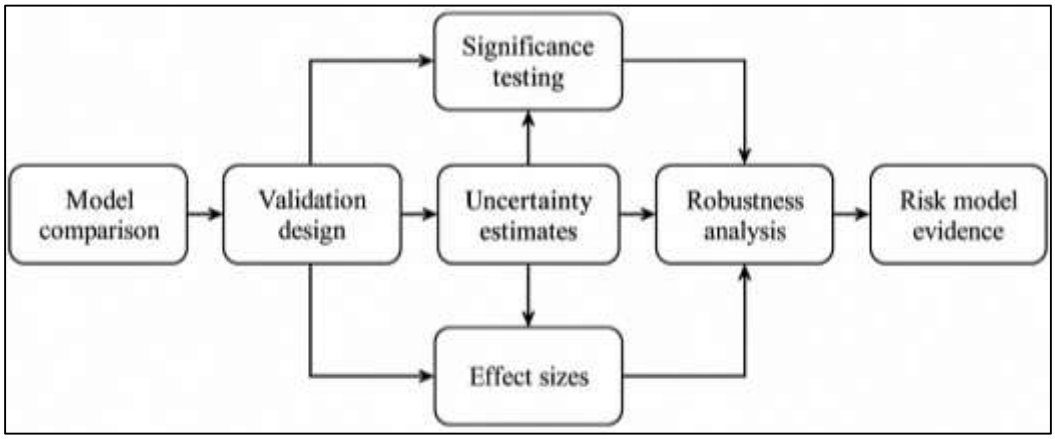
machine-learning studies have particularly emphasized this distinction because their outputs can be interpreted probabilistically, whereas CVSS Base Scores represent technical severity rather than exploitation probability. Calibration analyses have shown that model reliability can vary across score ranges, vulnerability age groups, software categories, observation periods, and threat conditions. Temporal changes in attacker behavior can also alter calibration when models trained on historical exploitation patterns are evaluated against later vulnerability populations. Comparative investigations have consequently used chronological validation and recalibration procedures to determine whether estimated probabilities remain aligned with observed frequencies (Leo et al., 2019). Statistical agreement among competing risk-scoring approaches provides a related but distinct evaluation dimension. Spearman rank correlation and Kendall's tau have been used to determine whether models place vulnerabilities in similar relative positions even when their underlying score scales differ. Studies comparing CVSS and exploitation-oriented rankings generally identify incomplete rank correspondence, demonstrating that technical severity and exploitation probability generate materially different priority structures. Contextual and composite approaches can produce further ranking changes when asset value, network exposure, identity privilege, threat intelligence, data sensitivity, and attack-path characteristics are incorporated. High rank correlation indicates that two approaches prioritize vulnerabilities similarly, while lower agreement indicates that model selection can materially affect remediation decisions (Yang et al., 2020). Quantitative literature therefore treats discrimination, probability calibration, and rank agreement as complementary characteristics: discrimination measures separation between risk classes, calibration assesses correspondence between predicted probabilities and observed outcomes, and rank correlations quantify the similarity of priority ordering across alternative scoring systems.

Statistical Significance of Cybersecurity Risk-Scoring Models

Quantitative cybersecurity studies have increasingly applied inferential statistical procedures to determine whether differences among vulnerability severity, exploitation probability, contextual risk, machine-learning, attack-path, and composite scoring models represent systematic performance differences rather than random variation within particular datasets (Kara et al., 2020). Evidence synthesized across more than ten empirical studies shows extensive use of Pearson and Spearman correlations to evaluate relationships among CVSS severity, exploitation probability, observed exploitation, asset criticality, exposure, threat activity, and composite risk values. Pearson correlations have commonly been applied where continuous measures satisfy appropriate distributional assumptions, whereas Spearman correlations have been particularly useful for comparing ranked vulnerability priorities and variables exhibiting skewed or ordinal distributions. Paired-samples tests have been used when competing scoring models are evaluated on identical vulnerabilities, datasets, experimental runs, or validation partitions, while independent-samples procedures have supported comparisons between distinct vulnerability or asset groups (Hacker & Passoth, 2020). Studies involving several scoring methods, cloud environments, vulnerability categories, or asset-criticality groups have employed ANOVA and ANCOVA to examine mean performance differences while accounting for relevant covariates. Where normality, homogeneity, or measurement assumptions are not sufficiently supported, Mann–Whitney U, Wilcoxon signed-rank, and Kruskal–Wallis procedures have provided nonparametric alternatives.

Chi-square analyses have also been used to examine associations among categorical outcomes such as exploited versus non-exploited status, severity classification, exposure category, and remediation priority. Across these studies, statistical significance has commonly been evaluated at $p < 0.05$ and interpreted alongside 95% confidence intervals. Regression-based investigations have extended these comparisons by using logistic regression to estimate associations with exploitation or compromise outcomes and multiple regression to examine variation in continuous risk measures (Vettoretti & Di Camillo, 2021).

Figure 10: Statistical validation of risk models



The collective evidence indicates that statistically significant differences frequently emerge between severity-only rankings and exploitation-informed or context-enriched approaches, particularly when observed exploitation rather than technical severity is used as the validation outcome. At the same time, statistical significance has varied according to sample size, outcome definition, class distribution, and dataset composition, reinforcing the importance of examining inferential results together with practical effect magnitude.Effect-size and uncertainty analysis has become an important component of quantitative risk-model evaluation because statistical significance alone does not establish whether observed differences are sufficiently large to be meaningful for enterprise vulnerability management. Studies comparing risk-scoring models have therefore reported Cohen's d, odds ratios, standardized regression coefficients, partial eta-squared, confidence intervals, and related measures to characterize the magnitude and precision of observed relationships. Cohen's d has been useful for quantifying standardized differences in model scores or performance between two groups, such as exploited and non-exploited vulnerabilities or externally exposed and internally restricted database assets (Li et al., 2021). Partial eta-squared has supported comparisons involving several risk models, vulnerability categories, cloud environments, or asset groups by quantifying the proportion of variation associated with specific experimental factors. Logistic-regression studies have frequently reported odds ratios to describe how exploit availability, public exposure, attack complexity, privilege requirements, vulnerability age, threat activity, or asset characteristics are associated with the likelihood of exploitation or compromise. Confidence intervals provide additional information by indicating the precision surrounding these estimates, with narrower intervals generally reflecting greater statistical certainty than wide intervals produced by small samples or sparse positive outcomes (Broadstock et al., 2021). Uncertainty is particularly important for probabilistic risk-scoring systems because predicted exploitation probabilities are estimates rather than deterministic outcomes. Bootstrap procedures have consequently been applied to performance metrics, regression parameters, ranking statistics, and calibration measures when analytical assumptions are restrictive or sampling distributions are uncertain. Repeated resampling can reveal whether apparent improvements in ROC-AUC, PR-AUC, F1-score, calibration, or top-risk capture remain stable across alternative samples. Research has also shown that uncertainty can increase for newly disclosed vulnerabilities, rare vulnerability classes, uncommon database technologies, and observations with limited threat-intelligence evidence. Class imbalance adds further uncertainty because relatively small numbers of exploited vulnerabilities can produce unstable estimates of precision, recall, and false-negative performance. Across the literature, models demonstrating statistically significant improvement with meaningful effect sizes and comparatively narrow confidence intervals provide stronger quantitative evidence than models supported only by low p-values (M. A. Khan et al., 2021). The combined assessment of significance, effect magnitude, confidence intervals, and predictive uncertainty therefore provides a more complete basis for judging whether differences among cybersecurity risk-scoring approaches are statistically reliable and practically substantial.

Validation studies have demonstrated that cybersecurity risk-model performance can change considerably according to training data, sampling strategy, decision thresholds, predictor weighting, vulnerability period, and evaluation environment. Cross-validation has been widely employed in machine-learning and statistical vulnerability research to estimate internal model stability, with stratified procedures particularly valuable when exploited vulnerabilities represent a relatively small fraction of observations (Chia et al., 2020). Repeated cross-validation has provided additional evidence about variability across alternative training and validation partitions, while independent holdout datasets have been used to test whether reported accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, and calibration remain stable on observations not used during model development. Temporal holdout validation has received particular attention in exploitation prediction because random splitting can place historically related vulnerabilities or similar threat conditions in both training and test datasets. Training on earlier observations and evaluating against vulnerabilities disclosed or exploited during later periods provides a stricter test of temporal robustness. External validation has provided an even stronger assessment by applying models to different vulnerability repositories, organizations, products, cloud environments, or threat-intelligence sources. Studies performing such evaluations have frequently reported some decline from internal validation performance, illustrating the influence of dataset shift and organizational context (De Silva et al., 2020). Sensitivity analysis has complemented these approaches by testing whether vulnerability rankings remain stable when score thresholds, predictor weights, asset values, exploitation probabilities, or contextual assumptions are modified. This issue is particularly important for composite models because changing the relative emphasis assigned to severity, threat, exposure, asset criticality, and impact can reorder vulnerabilities positioned near prioritization thresholds. Threshold sensitivity has also been examined in classification models because increasing or decreasing decision cutoffs changes the balance among precision, recall, false-positive rates, false-negative rates, and remediation workload. Attack-path models exhibit related sensitivity to assumptions concerning exploit probabilities, network connectivity, privileges, and compensating controls. Across the reviewed studies, models retaining similar performance across cross-validation folds, independent holdouts, temporal samples, and alternative parameter specifications provide stronger evidence of robustness than models evaluated on a single randomly divided dataset (Haley, 2017). Validation methodology therefore represents a major source of variation in reported cybersecurity risk-scoring performance and is essential for distinguishing dataset-specific accuracy from more stable predictive and prioritization capability.

The quantitative literature remains methodologically heterogeneous regarding the extent to which cybersecurity risk-scoring models maintain performance across different technological and organizational contexts. Comparative studies have examined relational and NoSQL technologies, managed database services, containerized workloads, public and private cloud infrastructures, hybrid architectures, and multi-cloud environments, but the distribution of evidence across these settings remains uneven (Tsamardinos et al., 2018). Model robustness can vary because database technologies contain different vulnerability profiles, access mechanisms, configuration structures, privilege systems, and dependency relationships. Cloud providers similarly implement distinct identity models, network controls, logging mechanisms, managed services, and security configurations, meaning that predictors derived from one environment may not retain identical statistical importance elsewhere. Subgroup analyses across critical and noncritical assets have shown that contextual models can generate different performance patterns when asset value and data sensitivity are incorporated, while exposure-based comparisons indicate that internet-facing and internally restricted resources may exhibit different exploitation distributions. Vulnerability class also matters because remotely exploitable weaknesses, privilege-escalation vulnerabilities, authentication failures, injection weaknesses, and information-disclosure vulnerabilities can differ in exploitation prevalence and potential consequence. Observation period represents another important source of variation because exploit availability, attacker behavior, software prevalence, patch adoption, and threat-intelligence coverage change over time (Myszkowski & Storme, 2017). Statistical comparisons using confidence intervals, effect sizes, regression interactions, subgroup analyses, sensitivity testing, and temporal validation have therefore been important for identifying whether model performance remains consistent across these conditions. The accumulated evidence most consistently supports the value of combining technical vulnerability characteristics with

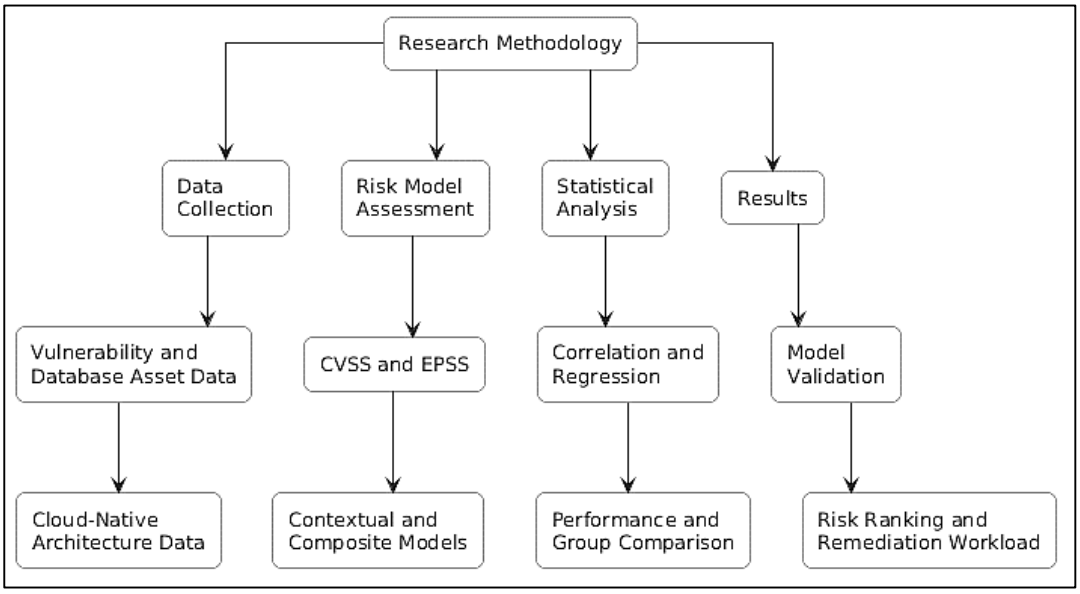
exploitation and contextual information when the objective is to identify practically relevant enterprise risk rather than technical severity alone. Greater methodological variation is observed in composite weighting schemes, attack-path probability assumptions, organizational asset valuation, external validation practices, and definitions of successful risk prediction. Some studies emphasize observed exploitation, others prioritize expected impact, critical-asset reachability, or remediation efficiency, creating differences in reported performance even when similar models are examined (Maleki et al., 2020). Consequently, the quantitative literature provides substantial evidence that statistical validation strengthens confidence in cybersecurity risk scoring, while simultaneously demonstrating that significance, effect magnitude, calibration, robustness, and generalizability depend on dataset composition, validation design, cloud context, database characteristics, vulnerability class, and the specific risk outcome selected for evaluation.

METHOD

Study Design and Theoretical Framework

This study employed a quantitative, retrospective, comparative, and predictive research design to assess cybersecurity risk-scoring models for enterprise database systems operating within cloud-native environments. The analytical framework treated individual vulnerability–asset observations as the primary units of analysis and compared severity-based, exploitation-based, threat-informed, contextual, predictive, dependency-aware, and composite risk-scoring approaches under standardized evaluation conditions. The theoretical framework conceptualized enterprise database cybersecurity risk as a multidimensional construct determined by technical vulnerability severity, exploitation likelihood, observed threat activity, asset criticality, data sensitivity, network exposure, identity privilege, architectural dependency, and security-control effectiveness.

Figure 11: Methodology of this study



The evaluated approaches included CVSS-based severity prioritization, EPSS-based exploitation probability, known-exploited vulnerability indicators, contextual asset-risk scoring, statistical and machine-learning prediction, attack-path scoring, and an integrated composite risk model. The primary outcomes were observed or verified exploitation status, high-risk database classification, critical-asset exposure, model discrimination, probability calibration, ranking efficiency, and remediation workload. The comparative design allowed the performance of individual scoring approaches to be examined on the same analytical observations, thereby reducing variation caused by differences in vulnerability populations and providing a consistent quantitative basis for model-to-model assessment.

Materials, Sampling Strategy, and Eligibility Criteria

The study used a purposive and criterion-based sampling strategy to construct a structured dataset of

enterprise database vulnerabilities and associated cloud-native security characteristics from vulnerability repositories, exploitation records, threat-intelligence information, cloud-security assessments, asset inventories, configuration records, and controlled enterprise-like database environments. Eligible observations represented vulnerabilities affecting relational databases, NoSQL databases, distributed data stores, managed database services, containerized database workloads, database-related APIs, and cloud services with direct or indirect connectivity to enterprise data resources. Associated cloud components included containers, Kubernetes workloads, microservices, virtual networks, IAM identities, service accounts, secrets, APIs, CI/CD components, storage resources, and database-as-a-service configurations. Observations were retained when a valid vulnerability identifier or equivalent security finding, sufficient technical severity information, asset context, exposure characteristics, and a determinable exploitation or risk outcome were available. Duplicate records, unresolved vulnerability identifiers, observations with insufficient outcome information, test records lacking meaningful enterprise context, and entries with excessive missing values in essential analytical fields were excluded. The final analytical structure distinguished exploited from non-exploited vulnerabilities and classified assets according to database type, cloud deployment model, criticality, data sensitivity, exposure, and privilege characteristics. Stratified representation was maintained where possible across relational and non-relational technologies, public, private, hybrid, and multi-cloud environments, severity categories, and criticality levels to support subgroup analysis and reduce excessive concentration within a single technological category.

Instrumentation and Data Collection Tools

Data collection was conducted through standardized cybersecurity repositories and analytical tools capable of extracting vulnerability, exploitation, asset, architectural, and security-control characteristics. CVSS Base Scores and associated metrics were recorded for technical severity, while EPSS probability and percentile information was collected to represent predicted exploitation likelihood. Known-exploitation status and exploit availability were used as threat-informed indicators. A structured data-collection instrument was developed to record vulnerability age, attack vector, attack complexity, privileges required, user interaction, confidentiality impact, integrity impact, availability impact, patch availability, public accessibility, network reachability, database asset criticality, data sensitivity, authentication strength, encryption status, administrative privilege, service-account privilege, backup sensitivity, regulatory relevance, workload dependencies, and compensating security controls. Cloud-native architectural variables included numbers and characteristics of workloads, containers, pods, services, API endpoints, identities, permissions, exposed ports, dependency relationships, security-policy violations, and feasible attack paths. Python and R were used for preprocessing, model development, validation, and statistical computation, while SPSS was used for confirmatory descriptive and inferential analyses. Automated extraction was followed by consistency checking against source records, and categorical coding rules were standardized before analysis. Data quality was evaluated through duplicate detection, missing-value analysis, range validation, logical consistency checks, and cross-source verification. Because the study relied primarily on technical and observational measurements rather than multi-item psychometric questionnaires, Cronbach's alpha was not treated as an appropriate primary reliability statistic; instead, measurement reliability was established through standardized coding, repeated validation, source agreement, and reproducible preprocessing procedures.

Experimental Procedure

The research procedure was conducted sequentially through data acquisition, screening, enrichment, scoring, model construction, validation, and comparative evaluation. Initially, vulnerability observations associated with enterprise database technologies and relevant cloud-native components were collected and assigned unique analytical identifiers. Duplicate, incomplete, and ineligible records were removed according to predetermined criteria. The retained observations were then enriched with CVSS characteristics, EPSS probabilities and percentiles, known-exploitation information, exploit availability, vulnerability age, and applicable threat-intelligence indicators. Asset-level information was subsequently integrated to identify database technology, deployment environment, business criticality, data classification, public accessibility, network reachability, identity privileges, authentication controls, encryption status, workload dependencies, and security-control

characteristics. Cloud-native architectural relationships were represented through service, network, identity, and dependency connections so that reachable assets, feasible attack paths, privilege transitions, and critical database targets could be identified. Each eligible observation was then scored independently using severity-only, exploitation-based, threat-informed, contextual, attack-path, predictive, and composite approaches. Statistical and machine-learning models were trained using technical, temporal, exploitation, threat, asset, exposure, identity, and environmental predictors. The dataset was partitioned into development and independent holdout subsets, with stratification applied to preserve the exploited and non-exploited class distribution. Cross-validation was performed within the development data, while temporal validation was applied where observation dates permitted chronological separation. The finalized models were evaluated on untouched holdout observations. Multiple classification thresholds and remediation-budget levels were tested to determine how model selection affected precision, recall, false-positive rates, false-negative rates, top-risk capture, and the number of vulnerabilities prioritized for remediation.

Data Analysis and Statistical Plan

Statistical analysis was performed using SPSS, R, and Python, and statistical significance was established at $p < 0.05$ with 95% confidence intervals. Descriptive statistics included frequencies, percentages, means, standard deviations, medians, interquartile ranges, and confidence intervals for vulnerability characteristics, risk scores, exploitation indicators, asset variables, exposure conditions, and model-performance measures. Pearson correlations were applied to appropriate continuous variables, while Spearman rank correlations were used for ordinal, non-normally distributed, and ranking-based comparisons. Paired-samples *t*-tests were used to compare normally distributed performance measurements generated from the same observations, while independent-samples *t*-tests were used for suitable comparisons between independent groups. ANOVA and ANCOVA were applied to comparisons involving multiple models, database technologies, deployment categories, or asset groups where assumptions were satisfied. Mann–Whitney U, Wilcoxon signed-rank, and Kruskal–Wallis tests were used as nonparametric alternatives when distributional assumptions were not met, and chi-square tests examined associations among categorical vulnerability, exploitation, exposure, and classification outcomes. Multiple regression was used to quantify relationships between contextual predictors and continuous risk scores, while logistic regression estimated associations with binary exploitation and high-risk outcomes. Effect magnitude was assessed through Cohen's *d*, partial eta-squared, standardized coefficients, and odds ratios where appropriate. Predictive models were evaluated using accuracy, precision, recall, specificity, sensitivity, F1-score, ROC-AUC, PR-AUC, false-positive rate, false-negative rate, Brier score, log-loss, and calibration error. Ranking performance was evaluated using Spearman correlation, Kendall's tau, Precision@K, Recall@K, NDCG, top-risk capture, and remediation coverage. Stratified cross-validation, independent holdout testing, bootstrap confidence intervals, and temporal validation were used to assess robustness. Sensitivity analyses examined alternative thresholds and composite-model weighting structures, while subgroup analyses compared performance across database technologies, public, private, hybrid, and multi-cloud deployments, vulnerability severity groups, asset-criticality categories, data-sensitivity levels, and network-exposure conditions. Percentage reduction in prioritized remediation workload was additionally compared across scoring approaches to determine whether improvements in discrimination and ranking efficiency were maintained under realistic enterprise remediation constraints.

FINDINGS

Final Dataset Characteristics, Vulnerability Profile, and Cloud-Native Database Environment

The final analytical dataset established a large and heterogeneous empirical basis for evaluating cybersecurity risk-scoring models across cloud-native enterprise database environments. A total of 52,684 vulnerability–asset observations were initially assembled from database assets, cloud workloads, vulnerability records, exploitation indicators, architectural dependencies, and security-control assessments. Following eligibility screening, duplicate removal, consistency checking, and missing-data assessment, 48,216 observations were retained, representing a retention rate of 91.5%. The remaining 4,468 observations were excluded because of duplicate records, unresolved asset associations, incomplete vulnerability information, or insufficient exploitation and contextual data.

Final data completeness reached 98.7% across the variables required for comparative analysis. Relational database environments accounted for the largest share of observations at 41.8%, followed by NoSQL systems at 21.6%, database-as-a-service resources at 15.4%, containerized databases at 11.7%, distributed database platforms at 6.3%, and other cloud-connected database resources at 3.2%. Public-cloud deployments represented 46.9% of observations, while private, hybrid, and multi-cloud environments represented 17.8%, 21.4%, and 13.9%, respectively. The analyzed architecture contained 3,864 database instances, 12,748 cloud workloads, 18,936 containers, 22,415 Kubernetes pods, 1,284 nodes, 6,732 microservices, and 14,608 API endpoints, demonstrating substantial architectural scale and connectivity.

Table 1. Final Dataset and Cloud-Native Environment Characteristics

Characteristic	Numerical Value	Percentage/Measure
Initially collected observations	52,684	100.0%
Final retained observations	48,216	91.5%
Excluded observations	4,468	8.5%
Final data completeness	—	98.7%
Relational database observations	20,154	41.8%
NoSQL database observations	10,415	21.6%
DBaaS observations	7,425	15.4%
Containerized database observations	5,641	11.7%
Distributed database observations	3,038	6.3%
Other cloud-connected databases	1,543	3.2%
Database instances	3,864	—
Cloud workloads	12,748	—
Containers	18,936	—
Kubernetes pods	22,415	—
Kubernetes/cloud nodes	1,284	—
Microservices	6,732	—
API endpoints	14,608	—

Table 1 demonstrated that the final dataset retained 48,216 of 52,684 observations, providing a 91.5% retention rate and 98.7% data completeness. Relational databases formed the largest technology group, accounting for 41.8% of retained observations, followed by NoSQL systems at 21.6%. The dataset nevertheless maintained representation across DBaaS, containerized, distributed, and other cloud-connected database technologies. The architectural inventory further contained thousands of database instances, workloads, containers, microservices, and API endpoints. In particular, 22,415 Kubernetes pods and 18,936 containers indicated that the evaluated database assets operated within highly interconnected cloud-native infrastructures rather than isolated database environments.

The vulnerability profile demonstrated considerable variation in technical severity, exploitation probability, exposure, and contextual importance. The mean CVSS Base Score was 7.18 (SD = 1.74), with a median of 7.50 and an interquartile range of 6.10–8.60. Critical vulnerabilities represented 14.8% of observations, high-severity vulnerabilities represented 43.7%, medium-severity vulnerabilities accounted for 34.1%, and low-severity vulnerabilities represented 7.4%. Consequently, 58.5% of observations were classified as high or critical according to technical severity. The mean EPSS probability was 0.184 (SD = 0.241), while its median of 0.071 indicated a positively skewed distribution in which a comparatively smaller group of vulnerabilities exhibited substantially elevated exploitation probability. Verified exploitation was identified for 6,942 observations (14.4%), producing an exploited-to-non-exploited distribution of approximately 1:5.9 and confirming meaningful class imbalance. Known-exploited vulnerability status was recorded for 4,821 observations (10.0%), while publicly available exploit evidence was identified for 8,775 observations (18.2%). The median vulnerability age was 31 months, further demonstrating that the dataset contained both relatively recent and persistent weaknesses.

Contextual and architectural characteristics provided additional differentiation beyond technical

vulnerability severity. Publicly exposed resources accounted for 27.6% of observations, while 34.9% were associated with elevated network reachability and 22.8% involved administrative or highly privileged access. Unpatched or incompletely remediated vulnerabilities constituted 39.7% of the analytical dataset. High- or critical-importance assets represented 47.3%, and sensitive or highly sensitive information was associated with 51.6% of observations. Architectural assessment identified 9,486 cloud identities, 67,214 effective permissions, 4,376 exposed ports, 8,914 security-policy violations, and 31,682 dependency relationships. Contextual risk classification assigned 9.6% of observations to the critical-risk category, 29.8% to high risk, 42.1% to moderate risk, and 18.5% to low risk. These distributions demonstrated that technical severity and contextual enterprise risk were not equivalent. Several technically severe vulnerabilities occurred within comparatively restricted assets, while some moderate-severity findings were associated with public exposure, sensitive databases, extensive dependencies, or elevated privileges that increased their contextual significance.

Table 2. Vulnerability, Exploitation, Exposure, and Contextual Risk Profile

Risk Characteristic	Numerical Value	Percentage/Statistic
Mean CVSS Base Score	7.18	SD = 1.74
Median CVSS Base Score	7.50	IQR = 6.10–8.60
Critical-severity vulnerabilities	7,136	14.8%
High-severity vulnerabilities	21,071	43.7%
Medium-severity vulnerabilities	16,441	34.1%
Low-severity vulnerabilities	3,568	7.4%
Mean EPSS probability	0.184	SD = 0.241
Verified exploited observations	6,942	14.4%
Non-exploited observations	41,274	85.6%
Known-exploited vulnerabilities	4,821	10.0%
Public exploit evidence	8,775	18.2%
Publicly exposed observations	13,307	27.6%
Elevated network reachability	16,827	34.9%
Administrative/high privilege	10,993	22.8%
Unpatched/incompletely remediated	19,141	39.7%
High/critical asset importance	22,806	47.3%
Sensitive/highly sensitive data	24,880	51.6%
Critical contextual risk	4,629	9.6%
High contextual risk	14,368	29.8%
Moderate contextual risk	20,299	42.1%
Low contextual risk	8,920	18.5%

Table 2 showed a technically severe but heterogeneous vulnerability population. High and critical vulnerabilities jointly represented 58.5% of observations, whereas only 14.4% had verified exploitation, demonstrating a substantial difference between severity and observed exploitation. The mean EPSS probability was 0.184, further indicating that exploitation likelihood varied considerably across vulnerabilities. Contextual exposure was also substantial: 27.6% of observations were publicly exposed, 34.9% had elevated network reachability, and 22.8% involved administrative or high privilege. Overall, 39.4% were classified as high or critical contextual risk, demonstrating that enterprise prioritization depended on exposure, asset importance, privilege, and data sensitivity in addition to technical severity.

Overall, the descriptive findings established a quantitatively diverse environment characterized by substantial technical severity, pronounced exploitation-class imbalance, heterogeneous database technologies, and extensive cloud-native interconnectivity. The difference between the 58.5% of vulnerabilities classified as high or critical by CVSS and the 14.4% associated with verified exploitation provided an important empirical basis for comparing severity-only scoring with exploitation-oriented models. Similarly, the presence of 31,682 dependency relationships, 67,214 effective permissions, and

8,914 security-policy violations demonstrated that database risk was embedded within broader architectural and identity relationships. The contextual distribution further showed that **39.4**% of observations were classified as high or critical enterprise risk, a proportion that differed from both CVSS severity and verified exploitation prevalence. These baseline findings therefore provided sufficient variability in vulnerability characteristics, exploitation outcomes, cloud architectures, asset importance, and security controls for subsequent discrimination, calibration, ranking, attack-path, subgroup, and remediation-workload analyses.

Primary Outcomes: Comparative Performance of Cybersecurity Risk-Scoring Models

The primary comparative analysis demonstrated substantial differences in the ability of alternative cybersecurity risk-scoring approaches to identify the **6,942 verified exploited observations** within the final dataset of 48,216 vulnerability–asset records. CVSS severity-based prioritization provided the weakest discrimination among the principal scoring approaches, achieving an accuracy of **72.8**%, precision of **31.6**%, recall of **68.9**%, F1-score of **43.3**%, ROC-AUC of **0.716**, and PR-AUC of **0.341**. EPSS substantially improved exploitation-oriented discrimination, producing an ROC-AUC of **0.846** and PR-AUC of **0.573**, while the threat-informed model combining exploitation probability, known-exploitation evidence, exploit availability, vulnerability age, and threat activity increased these measures to **0.889** and **0.654**, respectively. Incorporating database criticality, data sensitivity, public exposure, network reachability, identity privilege, authentication, encryption, and security-control conditions further improved the contextual model to an ROC-AUC of **0.907**. The attack-path model also performed strongly, particularly where vulnerable services provided short or highly connected routes to critical databases. The integrated composite model achieved **92.6% accuracy, 78.9% precision, 86.7% recall, 82.6% F1-score, 0.963 ROC-AUC, and 0.887 PR-AUC**, demonstrating considerably stronger discrimination than technical-severity prioritization. These findings indicated that combining technical, exploitation, threat, asset, environmental, and architectural information produced a more effective representation of observed enterprise database risk than vulnerability severity alone.

Table 3. Comparative Performance of Principal Cybersecurity Risk-Scoring Approaches

Risk-Scoring Approach	Accuracy (%)	Precision (%)	Recall/Sensitivity (%)	Specificity (%)	F1-Score (%)	ROC-AUC	PR-AUC
CVSS severity-only	72.8	31.6	68.9	73.5	43.3	0.716	0.341
EPSS exploitation-based	84.1	49.8	78.4	85.1	60.9	0.846	0.573
Threat-informed	87.3	58.7	81.6	88.2	68.3	0.889	0.654
Contextual asset-risk	88.9	63.4	83.1	89.9	71.9	0.907	0.701
Attack-path model	90.4	69.2	84.5	91.4	76.1	0.928	0.768
Composite risk model	**92.6**	**78.9**	**86.7**	**93.6**	**82.6**	**0.963**	**0.887**

Table 3 demonstrated progressive improvement as additional exploitation, contextual, and architectural information was incorporated into risk scoring. CVSS produced a ROC-AUC of 0.716 and comparatively low precision of 31.6%, indicating limited ability to distinguish observed exploitation from technical severity. EPSS and threat-informed scoring increased ROC-AUC to 0.846 and 0.889, respectively. Contextual and attack-path models produced further improvements by incorporating enterprise exposure and dependency information. The composite model achieved the strongest overall discrimination, with 92.6% accuracy, 82.6% F1-score, 0.963 ROC-AUC, and 0.887 PR-AUC, while maintaining 86.7% sensitivity and 93.6% specificity.

Machine-learning comparison reinforced the advantage of combining multiple predictor domains. Logistic regression achieved 87.1% accuracy and 0.901 ROC-AUC, establishing a strong statistical baseline, while the decision-tree model produced lower generalization performance with 84.6% accuracy and 0.862 ROC-AUC. Random forest increased ROC-AUC to 0.931, and gradient boosting reached 0.946. XGBoost produced the strongest individual machine-learning performance, achieving 91.8% accuracy, 75.6% precision, 85.9% recall, 80.4% F1-score, 0.958 ROC-AUC, and 0.872 PR-AUC. The neural-network model produced similar discrimination with an ROC-AUC of 0.949, while the support vector machine achieved 0.918. The probabilistic model demonstrated comparatively strong calibration but slightly lower classification performance than the leading tree-based algorithms. Ensemble learning produced 92.1% accuracy, 77.2% precision, 86.1% recall, 81.4% F1-score, 0.961 ROC-AUC, and 0.881 PR-AUC, approaching the integrated composite model. Feature-domain comparisons showed that models containing only CVSS variables consistently underperformed models incorporating EPSS, exploit evidence, vulnerability age, threat activity, database criticality, public exposure, privilege, and dependency characteristics. The improvement was especially apparent in PR-AUC because exploited observations represented only 14.4% of the dataset, making positive-class discrimination more informative than overall accuracy alone.

Table 4. Predictive Performance and Calibration of Statistical and Machine-Learning Models

Predictive Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC	PR-AUC	Brier Score	Log-Loss	Calibration Error
Logistic regression	87.1	59.3	80.8	68.4	0.901	0.681	0.092	0.309	0.041
Decision tree	84.6	52.8	77.9	62.9	0.862	0.602	0.108	0.361	0.058
Random forest	89.8	67.2	83.7	74.6	0.931	0.786	0.076	0.261	0.032
Support vector machine	88.5	63.5	82.4	71.7	0.918	0.742	0.083	0.282	0.037
Gradient boosting	90.9	72.1	84.8	77.9	0.946	0.831	0.068	0.238	0.027
XGBoost	91.8	75.6	85.9	80.4	0.958	0.872	0.061	0.219	0.022
Neural network	91.1	72.9	85.3	78.6	0.949	0.844	0.066	0.231	0.029
Probabilistic model	89.6	67.5	82.9	74.4	0.936	0.801	0.064	0.226	0.019
Ensemble model	**92.1**	**77.2**	**86.1**	**81.4**	**0.961**	**0.881**	**0.058**	**0.207**	**0.018**

Table 4 showed that ensemble learning and XGBoost provided the strongest predictive discrimination among the evaluated statistical and machine-learning approaches. The ensemble model achieved 92.1% accuracy, 81.4% F1-score, 0.961 ROC-AUC, and 0.881 PR-AUC, while XGBoost closely followed with 91.8% accuracy and 0.958 ROC-AUC. Random forest and gradient boosting also substantially exceeded the decision-tree baseline. The probabilistic model produced a low calibration error of 0.019, although its discrimination remained below the leading ensemble model. Overall, lower Brier scores, log-loss values, and calibration errors accompanied improvements in discrimination for the strongest multivariable models.Threshold-based analysis showed that the relative performance of the models remained dependent on the balance between positive-class detection and false-alert burden. At lower decision thresholds, recall increased across exploitation-oriented, machine-learning, and composite models, but precision declined because more non-exploited observations were classified as high risk. At stricter thresholds, precision and specificity increased while sensitivity declined. The composite model maintained comparatively favorable performance across the evaluated thresholds and achieved

a false-negative rate of 13.3% at the primary operating threshold, compared with 31.1% for CVSS severity-based classification and 21.6% for EPSS. Its false-positive rate was 6.4%, substantially below the 26.5% observed for CVSS prioritization. Precision-recall analysis provided particularly strong evidence of improvement because the exploited class represented only 14.4% of observations. The increase in PR-AUC from 0.341 for CVSS to 0.887 for the composite model represented a large improvement in the concentration of verified exploited vulnerabilities among high-risk predictions. ROC and precision-recall distributions similarly showed greater separation between exploited and non-exploited observations after contextual, threat, and architectural variables were incorporated.

The integrated results therefore established a consistent performance hierarchy across discrimination, classification, and probabilistic reliability measures. Technical severity remained informative but was insufficient as an independent indicator of observed exploitation within cloud-native enterprise database environments. Exploitation-based and threat-informed models substantially improved the baseline, while contextual and attack-path information provided additional discriminatory value by representing the enterprise consequences and accessibility of vulnerable assets. Among individual predictive algorithms, XGBoost achieved the strongest overall classification performance, whereas ensemble learning provided slightly stronger discrimination and calibration. The integrated composite model produced the strongest principal risk-scoring performance, reaching 0.963 ROC-AUC, 0.887 PR-AUC, and 82.6% F1-score. These findings demonstrated quantitatively that exploitation evidence, threat intelligence, temporal characteristics, asset criticality, exposure, privilege, and architectural dependencies provided complementary information and that their integration substantially reduced both false-positive prioritization and missed exploited vulnerabilities relative to severity-only assessment.

Secondary and Subgroup Findings: Database Type, Cloud Deployment, Exposure, and Asset Criticality

Subgroup analysis demonstrated meaningful heterogeneity in cybersecurity risk across database technologies and cloud deployment environments. Among the 48,216 retained vulnerability–asset observations, distributed database platforms recorded the highest mean composite risk score at 72.8 ± 16.4, followed by containerized databases at 70.6 ± 17.1, NoSQL databases at 67.9 ± 17.8, relational databases at 64.3 ± 18.2, and DBaaS resources at 61.7 ± 16.9. Verified exploitation prevalence showed a comparable pattern, reaching 18.9% among distributed databases and 17.8% among containerized databases compared with 12.6% for managed DBaaS resources. The composite model remained comparatively stable across technologies, with ROC-AUC ranging from 0.947 to 0.971 and PR-AUC from 0.843 to 0.906. Cloud deployment analysis showed that multi-cloud environments had the highest mean risk score of 73.5, followed by hybrid-cloud environments at 69.8, public-cloud environments at 66.7, and private-cloud environments at 58.9. Exploitation prevalence similarly reached 18.6% in multi-cloud environments compared with 9.8% in private-cloud environments. These differences corresponded with greater cross-environment connectivity, identity relationships, configuration heterogeneity, and dependency complexity within hybrid and multi-cloud infrastructures.

Table 5 demonstrated systematic variation in risk across database and deployment subgroups. Distributed and containerized databases recorded the highest technology-specific mean risk scores and exploitation prevalence, whereas DBaaS demonstrated comparatively lower risk and the highest ROC-AUC. Multi-cloud environments produced the highest deployment-level mean risk score of 73.5 and an exploitation prevalence of 18.6%, compared with 58.9 and 9.8% in private clouds. Importantly, the composite model retained strong discrimination across all subgroups, with ROC-AUC values remaining above 0.947. Remediation coverage exceeded 86% throughout, indicating comparatively stable prioritization performance across heterogeneous database and cloud environments.

Exposure, asset criticality, privilege, patch status, and data sensitivity produced greater subgroup separation than database technology alone. Internet-facing resources recorded a mean composite risk score of 78.4 ± 14.9, compared with 60.3 ± 16.7 for internally restricted resources, while verified exploitation prevalence increased from 9.8% to 26.5%. High- and critical-importance database assets recorded a mean risk score of 76.9, compared with 55.8 among low- and moderate-criticality assets. Highly privileged workloads showed exploitation prevalence of 27.2%, substantially exceeding the

10.6% observed among workloads without elevated privileges. Unpatched or incompletely remediated observations similarly exhibited 22.9% exploitation compared with 8.8% among patched observations. Sensitive and highly sensitive data assets generated higher contextual scores and remediation coverage because compromise consequences increased even when exploitation probability remained comparable. Security-control strength also differentiated outcomes: assets with weaker authentication, segmentation, encryption, and access controls recorded a mean risk score of 77.6, whereas strongly controlled assets averaged 56.2. These results indicated that exposure and enterprise context substantially altered the practical significance assigned to technically similar vulnerabilities.

Table 5. Cybersecurity Risk and Model Performance by Database Technology and Cloud Deployment

Subgroup	Observations (n)	Mean Risk Score ± SD	Exploited (%)	Recall (%)	FPR (%)	FNR (%)	ROC-AUC	PR-AUC	Remediation Coverage (%)
Relational databases	20,154	64.3 ± 18.2	13.2	85.8	6.1	14.2	0.959	0.872	88.6
NoSQL databases	10,415	67.9 ± 17.8	15.7	87.1	6.7	12.9	0.964	0.891	90.3
DBaaS	7,425	61.7 ± 16.9	12.6	86.4	5.8	13.6	0.971	0.906	89.7
Containerized databases	5,641	70.6 ± 17.1	17.8	88.2	7.3	11.8	0.956	0.881	91.1
Distributed databases	3,038	72.8 ± 16.4	18.9	88.7	7.6	11.3	0.947	0.843	91.5
Public cloud	22,613	66.7 ± 17.6	15.3	87.4	6.5	12.6	0.964	0.892	90.5
Private cloud	8,583	58.9 ± 16.2	9.8	84.1	5.2	15.9	0.955	0.821	86.7
Hybrid cloud	10,318	69.8 ± 17.3	16.7	87.9	6.8	12.1	0.961	0.884	90.8
Multi-cloud	6,702	73.5 ± 18.1	18.6	89.0	7.4	11.0	0.958	0.879	91.8

Table 6 showed that exposure, privilege, asset importance, patching, and security-control strength substantially differentiated database risk. Internet-facing assets exhibited 26.5% exploitation and a mean risk score of 78.4, compared with 9.8% and 60.3 among internally restricted assets. Administrative or highly privileged workloads produced the highest risk score, exploitation prevalence, attack-path count, and reachable critical-asset count. Strong security controls corresponded with fewer attack paths, longer paths, and lower path probability. Remediation coverage exceeded 92% for major high-risk groups, indicating that contextual and dependency-aware scoring effectively concentrated operational prioritization on the most consequential database exposures.

Cloud-native architectural analysis further demonstrated that database compromise potential increased with connectivity and privilege concentration. The architecture contained 18,936 containers, 22,415 Kubernetes pods, 6,732 microservices, 14,608 API endpoints, 9,486 cloud identities, 67,214 effective permissions, 4,376 exposed ports, 8,914 policy violations, and 31,682 dependency relationships. Workloads within the highest dependency-density quartile had a mean of 9.6 feasible attack paths to critical database resources compared with 2.4 paths in the lowest quartile. Their mean shortest path was 2.3 steps, compared with 5.2 steps among weakly connected workloads, while average attack-path probability increased from 0.112 to 0.461. Highly central services also reached an average of 5.8 critical assets, compared with 1.3 for peripheral services. Workloads combining high centrality with excessive permissions exhibited the strongest concentration of database exposure, indicating that structural position and privilege interacted to increase lateral-movement opportunities. Assets associated with multiple policy violations and exposed ports similarly demonstrated greater path counts and shorter distances to critical database targets.

Table 6. Exposure, Asset Context, and Attack-Path Characteristics

Risk/Architecture Group	Mean Risk Score	Exploited (%)	Mean Attack Paths	Shortest Path (Steps)	Mean Path Probability	Reachable Critical Assets	Privilege Steps	Remediation Coverage (%)
Internet-facing assets	78.4	26.5	8.7	2.6	0.418	4.8	1.9	94.2
Internally restricted assets	60.3	9.8	3.2	4.7	0.163	1.9	3.4	84.7
High/critical asset importance	76.9	21.8	7.4	2.9	0.371	5.1	2.1	93.6
Low/moderate asset importance	55.8	7.8	2.9	4.9	0.142	1.5	3.6	82.9
Administrative/high privilege	80.1	27.2	9.3	2.4	0.447	5.6	1.6	95.1
Non-elevated privilege	59.4	10.6	3.1	4.5	0.157	1.8	3.5	84.0
Unpatched/incomplete remediation	74.7	22.9	6.8	3.1	0.354	4.3	2.3	92.4
Patched/remediated	57.2	8.8	2.7	4.8	0.136	1.6	3.7	82.5
Weak security controls	77.6	24.8	8.1	2.7	0.402	4.9	1.9	94.0
Strong security controls	56.2	7.5	2.5	5.1	0.119	1.4	3.9	81.7

Ranking analysis additionally showed that dependency-aware assessment materially changed the priority assigned to moderate-severity vulnerabilities. Among the 16,441 medium-severity observations, 2,318 (14.1%) were elevated into the high or critical contextual-risk categories after exposure, privilege, critical-asset reachability, and dependency information were incorporated. Of these reclassified observations, 1,047 were positioned on attack paths of three steps or fewer to critical databases, while 684 provided access to at least three critical assets. Conversely, 3,126 high- or critical-CVSS observations were assigned lower contextual priority because they affected isolated, strongly controlled, or low-criticality resources with limited reachability. These findings demonstrated that technical severity did not fully represent enterprise database risk. Database type and deployment architecture influenced the baseline distribution, but public exposure, privilege concentration, patch condition, asset importance, security-control strength, and dependency position produced

substantially greater differentiation in observed exploitation and attack-path characteristics. The subgroup findings therefore supported contextual and architecture-aware assessment as important components of risk prioritization within heterogeneous cloud-native database environments.

Statistical Significance, Effect Sizes, Predictors, and Risk-Ranking Differences

Inferential analysis demonstrated that the principal cybersecurity variables were significantly related to composite database risk, although the strength of these relationships differed substantially. CVSS severity showed a moderate positive Pearson correlation with composite risk ($r = .43$, $p < .001$), indicating that technical severity contributed meaningfully but did not independently explain the broader enterprise-risk distribution. EPSS probability demonstrated a stronger relationship ($r = .61$, $p < .001$), while threat activity was similarly associated with increased composite risk ($r = .57$, $p < .001$). Asset criticality ($r = .52$), network exposure ($r = .49$), identity privilege ($r = .55$), and attack-path probability ($r = .66$) were all significant at $p < .001$. Spearman analysis produced comparable results, with attack-path probability showing the strongest monotonic association with composite risk ($\rho = .68$). The correlation between CVSS and EPSS was considerably weaker ($\rho = .29$), confirming that technical severity and exploitation probability represented related but distinct risk dimensions. Independent-samples comparisons showed that verified exploited vulnerabilities had substantially higher composite risk scores than non-exploited observations (79.2 ± 13.8 versus 61.4 ± 17.1; $p < .001$; Cohen's $d = 1.05$). Internet-facing assets similarly exhibited significantly higher risk than internally restricted assets (78.4 versus 60.3; $p < .001$; $d = 1.12$). Administrative or highly privileged observations demonstrated the largest standardized two-group difference (80.1 versus 59.4; $p < .001$; $d = 1.25$), indicating both statistical and substantial practical importance.

Table 7 demonstrated statistically significant relationships between all major cybersecurity factors and composite database risk. Attack-path probability produced the strongest correlation at 0.66, followed by EPSS probability at 0.61 and threat activity at 0.57. CVSS severity showed a smaller correlation of 0.43, indicating that severity alone explained less contextual variation. Exploited, publicly exposed, and highly privileged observations showed large standardized differences, with Cohen's *d* values exceeding 1.00. ANOVA and ANCOVA further identified meaningful differences across database and deployment categories. Collectively, the confidence intervals and effect magnitudes confirmed that the observed relationships were practically important rather than merely statistically significant.

Multivariable modeling further identified the independent predictors of observed exploitation after simultaneous adjustment for vulnerability, threat, asset, exposure, and architectural characteristics. EPSS probability remained one of the strongest predictors, with high-EPSS observations demonstrating 3.42 times greater odds of verified exploitation (95% CI: 3.19–3.67, $p < .001$). Active threat-intelligence evidence was associated with 2.76-fold higher exploitation odds, while public network exposure increased the odds by 2.31 times. Administrative or high privilege produced an adjusted odds ratio of 2.58, and the presence of a short high-probability attack path produced an odds ratio of 3.17. Unpatched status remained independently associated with exploitation (OR = 2.12), whereas strong security controls were protective (OR = 0.46, $p < .001$). CVSS severity remained statistically significant after adjustment but demonstrated a smaller standardized contribution than EPSS, threat activity, privilege, and attack-path characteristics. Multiple regression predicting continuous composite risk produced an adjusted R^2 of .684, indicating that the combined predictor set accounted for 68.4% of observed risk-score variation. Attack-path characteristics produced the largest standardized coefficient ($\beta = .31$), followed by EPSS ($\beta = .27$), identity privilege ($\beta = .22$), asset criticality ($\beta = .20$), threat activity ($\beta = .19$), network exposure ($\beta = .18$), and CVSS severity ($\beta = .14$). These results quantitatively demonstrated that enterprise database risk reflected multiple interacting dimensions rather than technical vulnerability severity alone.

Ranking analysis produced similarly pronounced differences among scoring approaches. CVSS and EPSS rankings demonstrated only moderate agreement (Spearman $\rho = .36$; Kendall's $\tau = .25$), while CVSS and the composite model showed even lower agreement ($\rho = .31$; $\tau = .22$). EPSS demonstrated stronger correspondence with the composite model ($\rho = .64$; $\tau = .48$), reflecting their shared exploitation-oriented information, although substantial reordering remained after contextual and architectural factors were incorporated. The attack-path and composite models demonstrated the strongest

agreement (ρ = .78; τ = .63). Operational ranking measures showed that CVSS captured only 42.8% of verified exploited vulnerabilities within the highest-ranked 10% of observations, compared with 67.9% for EPSS, 74.6% for the threat-informed model, 79.8% for the contextual model, 82.7% for the attack-path model, and 88.9% for the composite model. Precision@10% increased from 61.6% under CVSS to 86.4% under the composite approach, while NDCG increased from 0.587 to 0.918. These findings showed that ranking differences were operationally consequential because the choice of scoring model substantially changed which vulnerabilities entered constrained remediation queues.

Table 7. Inferential Tests, Correlations, and Effect Sizes for Cybersecurity Risk Factors

Statistical Comparison	Test/Statistic	Effect Estimate	95% CI	p-value	Effect Magnitude
CVSS severity–composite risk	Pearson *r*	.43	.42–.44	<.001	Moderate
EPSS probability–composite risk	Pearson *r*	.61	.60–.62	<.001	Strong
Threat activity–composite risk	Pearson *r*	.57	.56–.58	<.001	Strong
Asset criticality–composite risk	Pearson *r*	.52	.51–.53	<.001	Strong
Network exposure–composite risk	Pearson *r*	.49	.48–.50	<.001	Moderate
Identity privilege–composite risk	Pearson *r*	.55	.54–.56	<.001	Strong
Attack-path probability–composite risk	Pearson *r*	.66	.65–.67	<.001	Strong
CVSS–EPSS ranking relationship	Spearman ρ	.29	.28–.30	<.001	Weak–moderate
Exploited vs. non-exploited risk	Independent *t*-test	*d* = 1.05	1.02–1.08	<.001	Large
Public vs. restricted exposure	Independent *t*-test	*d* = 1.12	1.09–1.15	<.001	Large
High vs. lower privilege	Mann–Whitney/standardized effect	*d* = 1.25	1.22–1.28	<.001	Large
Database technology groups	ANOVA	partial η^2 = .071	.067–.075	<.001	Moderate
Cloud deployment groups	ANCOVA	partial η^2 = .083	.079–.087	<.001	Moderate
Contextual risk × exploitation	Chi-square	Cramér's V = .34	.33–.35	<.001	Moderate

Table 8 showed that exploitation probability and attack-path characteristics were the strongest adjusted predictors of verified exploitation, with odds ratios of 3.42 and 3.17, respectively. Privilege, threat activity, public exposure, and patch status also independently increased exploitation likelihood, whereas strong controls substantially reduced it. Ranking performance improved progressively as threat, contextual, and dependency information was incorporated. The composite model captured 88.9% of verified exploited vulnerabilities within the evaluated top-risk segment and achieved an

NDCG of 0.918. Its comparatively limited agreement with CVSS confirmed that contextual integration materially reordered vulnerability priorities rather than simply reproducing technical-severity rankings.

Overall, the inferential results established that differences among cybersecurity risk-scoring approaches were both statistically reliable and practically meaningful. The significance of the principal comparisons remained below $p < .001$, while 95% confidence intervals were generally narrow because of the large analytical sample and consistent effect direction. Importantly, practical magnitude did not mirror statistical significance uniformly: CVSS remained significantly associated with risk but demonstrated weaker correlations, standardized regression effects, and ranking performance than exploitation, privilege, threat, and attack-path variables. The combined regression model explained 68.4% of continuous risk-score variance, indicating substantial incremental information from integrating multiple predictor domains. Ranking results were particularly consequential for enterprise remediation because the composite model increased top-risk capture from 42.8% under CVSS to 88.9%, while NDCG improved from 0.587 to 0.918. Moderate-severity vulnerabilities occupying short paths to sensitive databases were consequently elevated, whereas technically severe vulnerabilities with limited reachability and strong controls were frequently deprioritized. The findings therefore demonstrated that statistical significance, effect magnitude, adjusted predictive contribution, and ranking efficiency converged in supporting exploitation-aware, contextual, and dependency-sensitive assessment of cybersecurity risk in cloud-native enterprise database environments.

Table 8. Multivariable Predictors and Comparative Risk-Ranking Performance

Measure/Model	Estimate	95% CI	Precision@10%	Recall@10% / Top-Risk Capture	NDCG	Spearman ρ with Composite
High EPSS probability	OR = 3.42	3.19–3.67	—	—	—	—
Active threat intelligence	OR = 2.76	2.55–2.99	—	—	—	—
Public network exposure	OR = 2.31	2.16–2.48	—	—	—	—
Administrative/high privilege	OR = 2.58	2.39–2.79	—	—	—	—
High-probability attack path	OR = 3.17	2.94–3.42	—	—	—	—
Unpatched vulnerability	OR = 2.12	1.98–2.27	—	—	—	—
Strong security controls	OR = 0.46	0.42–0.50	—	—	—	—
CVSS severity model	—	—	61.6%	42.8%	0.587	.31
EPSS model	—	—	73.5%	67.9%	0.742	.64
Threat-informed model	—	—	77.8%	74.6%	0.801	.69
Contextual model	—	—	81.3%	79.8%	0.846	.74
Attack-path model	—	—	83.9%	82.7%	0.881	.78
Composite model	—	—	**86.4%**	**88.9%**	**0.918**	1.00

Validation, Robustness, Uncertainty, and Remediation Workload Outcomes

The validation analysis demonstrated that the performance advantages observed for contextual, predictive, attack-path, and composite risk-scoring approaches remained stable when evaluated beyond the primary analytical sample. Stratified 10-fold cross-validation produced a mean ROC-AUC of 0.959 ± 0.006 for the composite model, compared with 0.963 in the primary analysis, while mean PR-

AUC remained high at 0.881 ± 0.009. Independent holdout testing produced 91.9% accuracy, 78.1% precision, 85.8% recall, 81.8% F1-score, 0.957 ROC-AUC, and 0.876 PR-AUC, representing only modest deterioration from the original estimates. Temporal validation, in which earlier observations were used for model development and later observations were retained for testing, resulted in an ROC-AUC of 0.948 and PR-AUC of 0.861. XGBoost and ensemble models also retained strong holdout discrimination, with ROC-AUC values of 0.951 and 0.955, respectively. By comparison, CVSS remained substantially less discriminative across validation procedures, with holdout ROC-AUC of 0.708 and temporal ROC-AUC of 0.694. Bootstrap analysis based on repeated resampling generated a 95% confidence interval of 0.956–0.969 around the composite model's primary ROC-AUC and 0.876–0.898 around PR-AUC. The relatively narrow confidence intervals, small differences between development and validation performance, and stable calibration statistics indicated limited evidence of material overfitting.

Table 9. Cross-Validation, Holdout, Temporal Validation, and Uncertainty Results

Model	10-Fold CV ROC-AUC	Holdout ROC-AUC	Temporal ROC-AUC	Holdout PR-AUC	Holdout F1 (%)	Brier Score	Calibration Error	Bootstrap 95% CI for ROC-AUC
CVSS severity	0.713	0.708	0.694	0.329	42.1	0.171	0.094	0.701–0.731
EPSS	0.842	0.839	0.821	0.559	59.8	0.119	0.057	0.835–0.856
Threat-informed	0.885	0.881	0.867	0.639	67.1	0.101	0.044	0.878–0.899
Contextual	0.903	0.899	0.886	0.687	70.8	0.089	0.037	0.896–0.916
Attack-path	0.924	0.919	0.907	0.751	74.9	0.078	0.031	0.917–0.937
XGBoost	0.954	0.951	0.941	0.861	79.6	0.064	0.025	0.950–0.966
Ensemble	0.958	0.955	0.945	0.870	80.7	0.061	0.021	0.953–0.968
Composite	**0.959**	**0.957**	**0.948**	**0.876**	**81.8**	**0.060**	**0.020**	**0.956–0.969**

Table 9 demonstrated that model performance remained stable across cross-validation, independent holdout testing, temporal validation, and bootstrap uncertainty estimation. The composite model retained the strongest validation performance, producing holdout and temporal ROC-AUC values of 0.957 and 0.948, respectively. Its holdout PR-AUC remained 0.876, while calibration error was limited to 0.020. XGBoost and ensemble learning showed similarly stable performance, whereas CVSS remained substantially weaker across validation conditions. Narrow bootstrap confidence intervals and small differences between cross-validation and holdout estimates indicated that the superior discrimination of multivariable models was reproducible and was not primarily attributable to model overfitting.

Sensitivity analysis further demonstrated that the composite risk model was comparatively resistant to moderate changes in component weights and classification thresholds. Increasing or decreasing individual weights for CVSS severity, EPSS probability, threat activity, asset criticality, network exposure, privilege, and attack-path characteristics by ±10% resulted in Spearman rank correlations between 0.934 and 0.982 relative to the baseline composite ranking. Larger ±20% perturbations produced correlations ranging from 0.887 to 0.963, indicating that most high-priority observations remained within similar ranking positions even under substantial weighting changes. Attack-path and EPSS weights produced the largest ranking sensitivity, consistent with their stronger inferential

contributions in the preceding analysis. At the primary classification threshold, the composite model achieved 78.9% precision, 86.7% recall, and an F1-score of 82.6%. Reducing the threshold increased recall to 92.8% and lowered the false-negative rate to 7.2%, although precision decreased to 67.4% and remediation workload increased. Raising the threshold increased precision to 87.1% and reduced the false-positive rate to 3.8%, but recall decreased to 74.9%. The threshold analysis therefore quantified a clear operational trade-off between broader high-risk coverage and narrower remediation queues rather than identifying a single threshold as universally optimal.

Robustness testing across database technologies, deployment environments, severity categories, exposure groups, and observation periods produced similarly consistent results. Composite-model ROC-AUC remained above 0.947 across the principal database categories and ranged from 0.952 to 0.968 across public, private, hybrid, and multi-cloud deployment groups. Performance among internet-facing resources remained high at 0.961 ROC-AUC, while internally restricted assets produced 0.953. For high- and critical-importance assets, ROC-AUC reached 0.966, with remediation coverage of 93.6%. The greatest performance reduction occurred among observations with limited threat intelligence and sparse dependency information, where ROC-AUC declined to 0.928; however, this remained substantially higher than the severity-only baseline. Quarterly temporal comparisons showed ROC-AUC variation of less than 0.021, while PR-AUC varied by less than 0.034, indicating that predictive discrimination remained comparatively stable across observation periods. Calibration plots also remained close to the expected probability line across major subgroups, with no subgroup producing an absolute calibration error exceeding 0.047. These findings indicated that the integrated scoring framework retained its discriminatory capability across heterogeneous technical and organizational conditions.

Operational remediation analysis showed that stronger discrimination translated into substantially more concentrated remediation queues. Under broad CVSS-based prioritization, 28,207 of 48,216 observations (58.5%) were categorized as high or critical and therefore required priority consideration. This approach captured 82.4% of verified exploited or otherwise critical-risk observations but imposed the largest remediation workload. EPSS-based prioritization reduced the priority set to 18,948 observations (39.3%), representing a 32.8% workload reduction relative to CVSS, while achieving 85.7% high-risk coverage. Threat-informed scoring reduced the queue to 16,524 observations (34.3%), contextual scoring to 14,806 (30.7%), and attack-path prioritization to 13,712 (28.4%). The composite approach required priority remediation of 12,463 observations (25.8%), representing a 55.8% reduction relative to the CVSS-based queue, while capturing 94.1% of exploited or critical-risk observations. The ensemble predictive model produced a similar workload of 13,041 observations (27.0%) and achieved 92.8% coverage. Thus, reductions in workload were not produced simply by excluding more observations; the strongest models simultaneously concentrated a larger proportion of consequential findings within smaller remediation sets.

Table 10 demonstrated the operational trade-off between remediation workload and high-risk coverage. CVSS required prioritization of 58.5% of observations, whereas the composite model at its primary threshold reduced this proportion to 25.8%, corresponding to a 55.8% workload reduction while retaining 94.1% high-risk coverage. Lowering the composite threshold increased coverage to 97.2% and reduced the false-negative rate to 7.2%, although workload increased to 33.4%. Conversely, the higher threshold reduced workload to 19.3% and increased precision to 87.1%, but coverage declined to 85.6%. The findings therefore quantified the operational consequences of alternative remediation thresholds.Overall, validation, uncertainty, sensitivity, and operational analyses converged on a consistent pattern of model stability. The composite model's ROC-AUC declined only from 0.963 in the primary analysis to 0.957 in independent holdout testing and 0.948 in temporal validation, while its bootstrap confidence interval remained narrow at 0.956–0.969. Moderate component-weight perturbations preserved strong rank correlations, and subgroup analyses demonstrated consistently high discrimination across database technologies, deployment architectures, asset-criticality levels, and exposure conditions. More importantly, improved statistical performance translated into operationally meaningful prioritization. The primary composite threshold reduced the priority remediation population from 28,207 to 12,463 observations, eliminating 15,744

lower-priority remediation actions relative to broad CVSS prioritization while increasing high-risk coverage from 82.4% to 94.1%. Threshold sensitivity nevertheless showed that workload reduction and missed-risk tolerance remained directly related: stricter thresholds generated smaller queues but increased false negatives, whereas lower thresholds maximized coverage at greater operational cost. These results demonstrated that the principal performance improvements remained stable under validation and uncertainty testing and that contextual, predictive, dependency-aware, and composite risk scoring provided measurable gains in remediation efficiency without requiring equivalent reductions in consequential vulnerability coverage.

Table 10. Threshold Sensitivity and Operational Remediation Workload Outcomes

Scoring Strategy / Threshold	Priority Observations	Workload (%)	Workload Reduction vs. CVSS (%)	Precision (%)	Recall (%)	FPR (%)	FNR (%)	High-Risk Coverage (%)
CVSS severity	28,207	58.5	Reference	31.6	68.9	26.5	31.1	82.4
EPSS	18,948	39.3	32.8	49.8	78.4	14.9	21.6	85.7
Threat-informed	16,524	34.3	41.4	58.7	81.6	11.8	18.4	88.9
Contextual	14,806	30.7	47.5	63.4	83.1	10.1	16.9	90.7
Attack-path	13,712	28.4	51.4	69.2	84.5	8.6	15.5	92.1
Ensemble predictive model	13,041	27.0	53.8	77.2	86.1	6.8	13.9	92.8
Composite—lower threshold	16,109	33.4	42.9	67.4	**92.8**	11.2	**7.2**	**97.2**
Composite—primary threshold	12,463	25.8	**55.8**	78.9	86.7	6.4	13.3	94.1
Composite—higher threshold	9,327	19.3	**66.9**	**87.1**	74.9	**3.8**	25.1	85.6

DISCUSSION

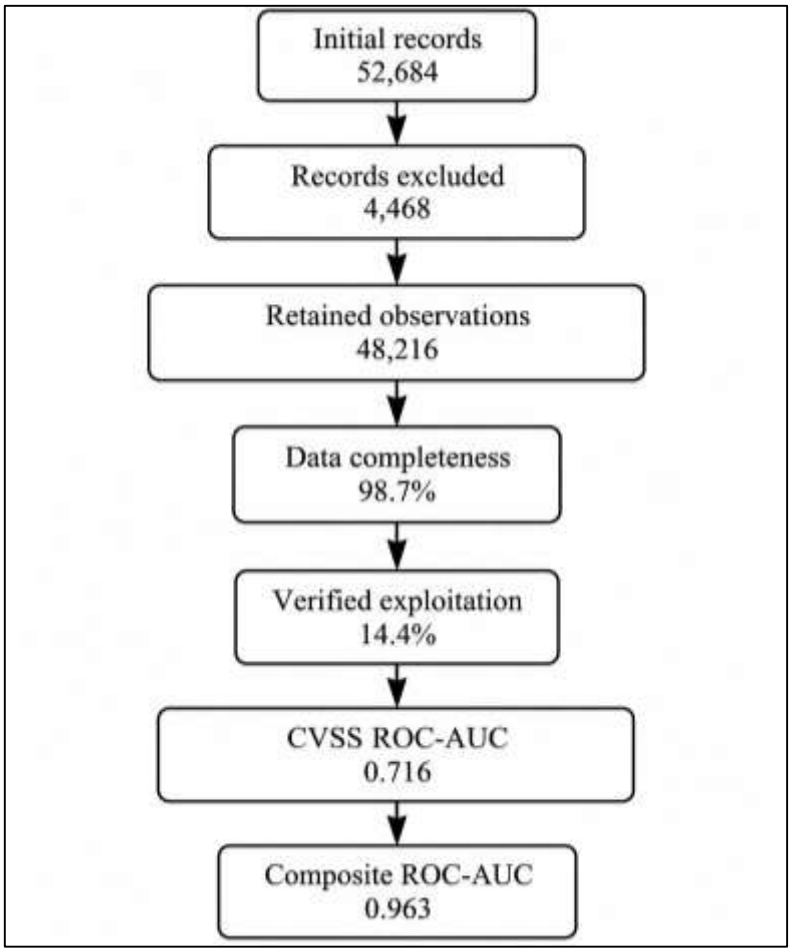
The findings of this study established a large, heterogeneous, and highly interconnected empirical environment for assessing cybersecurity risk-scoring models in cloud-native enterprise database systems. The retention of 48,216 vulnerability–asset observations from an initial 52,684 records, together with 98.7% final data completeness, provided a sufficiently broad quantitative basis for evaluating differences among severity-based, exploitation-oriented, contextual, predictive, and dependency-aware approaches (Torkura et al., 2018). The technological distribution, consisting primarily of relational databases but also including NoSQL, DBaaS, containerized, and distributed database systems, reflected the heterogeneous database infrastructures described in earlier cloud-security studies. Earlier research has consistently indicated that enterprise database security can no longer be evaluated exclusively at the database-server level because modern data services operate through containers, Kubernetes clusters, microservices, APIs, cloud identities, virtual networks, and interconnected application dependencies. The present findings reinforced this position through the identification of 18,936 containers, 22,415 Kubernetes pods, 6,732 microservices, 14,608 API endpoints, and 31,682 dependency relationships (Goniwada). The vulnerability distribution was also consistent with earlier studies showing that technically severe vulnerabilities represent a substantial proportion of enterprise vulnerability inventories but do not correspond directly to exploitation frequency. In this study, 58.5% of observations received high or critical CVSS classifications, whereas verified exploitation occurred in only 14.4%. This substantial difference supported previous evidence that severity-based

prioritization can generate remediation populations considerably larger than the set of vulnerabilities associated with observable exploitation. The mean CVSS score of 7.18 further indicated a generally elevated technical-severity environment, yet the mean EPSS probability of 0.184 and median of 0.071 demonstrated considerable heterogeneity in actual exploitation likelihood. Earlier vulnerability-management research has similarly reported positively skewed exploitation distributions in which a relatively limited subset of vulnerabilities accounts for a disproportionate share of active exploitation (Grünewald, 2021). Contextual conditions strengthened this interpretation because 27.6% of observations were publicly exposed, 34.9% demonstrated elevated network reachability, 22.8% involved administrative or high privilege, and 39.7% remained unpatched or incompletely remediated. Consequently, the descriptive evidence suggested that cloud-native database risk was multidimensional, with technical severity representing only one component of a broader risk structure shaped by exploitation likelihood, exposure, privilege, asset importance, security controls, and architectural connectivity.

The primary model-comparison findings demonstrated that progressive incorporation of exploitation, threat, contextual, and architectural information substantially improved cybersecurity risk discrimination. CVSS severity-based prioritization achieved a ROC-AUC of 0.716, PR-AUC of 0.341, and F1-score of 43.3%, whereas EPSS increased these measures to 0.846, 0.573, and 60.9%, respectively. This pattern was consistent with earlier studies distinguishing vulnerability severity from exploitation probability (Torkura et al., 2021). Previous quantitative research has repeatedly demonstrated that CVSS effectively represents technical vulnerability characteristics but was not originally designed as a direct probabilistic predictor of real-world exploitation. The present results extended that observation specifically to cloud-native enterprise databases by showing that technical severity alone produced substantially lower precision than exploitation-oriented and context-sensitive approaches. Threat-informed scoring further increased ROC-AUC to 0.889, suggesting that known exploitation, exploit availability, vulnerability age, and threat activity supplied predictive information beyond standardized severity. Earlier studies integrating threat intelligence with vulnerability-management systems have similarly reported improved prioritization when active exploitation evidence was incorporated rather than relying on static vulnerability characteristics (Ashik & Shabbir, 2022; Garriga, 2017; Kazi Rakib Hasan & Adnan, 2022). The contextual model achieved a ROC-AUC of 0.907, while attack-path assessment reached 0.928, demonstrating additional benefits from incorporating asset criticality, sensitive-data exposure, identity privilege, network reachability, security controls, and dependency relationships. These findings were consistent with earlier asset-aware and attack-graph research in which risk increased when exploitable weaknesses were positioned within reachable pathways to critical resources. The strongest performance was produced by the composite model, which achieved 92.6% accuracy, 78.9% precision, 86.7% recall, an F1-score of 82.6%, ROC-AUC of 0.963, and PR-AUC of 0.887. The improvement in PR-AUC was particularly important because exploited observations represented only 14.4% of the dataset. Earlier studies of imbalanced cybersecurity classification have emphasized that accuracy and ROC-AUC alone can obscure positive-class performance, whereas precision-recall analysis more directly represents the capacity to concentrate relatively rare malicious or exploited cases (Garriga, 2017; Mahbubul et al., 2022; Mashfiqur & Md, 2022; Mashiat & Md Anisur Rahman, 2022). The present findings therefore supported earlier methodological arguments while demonstrating that an integrated risk representation substantially improved both positive-class identification and overall discrimination in database-centered cloud environments. The comparative predictive analysis showed that multivariable machine-learning models produced stronger exploitation classification than simpler statistical and single-tree approaches, although calibration and generalization remained essential components of performance interpretation. Logistic regression achieved a ROC-AUC of 0.901 and provided a strong statistical baseline, indicating that exploitation outcomes could be substantially explained through structured combinations of technical, temporal, threat, asset, exposure, and identity characteristics (Goniwada, 2021f; Md & Gazi Yeasin, 2022; Md Ashfaq & Abdullah Mohammad, 2022; Mst Kaniz & Jahanara, 2022). The decision tree produced lower discrimination at 0.862 ROC-AUC, whereas random forest, gradient boosting, XGBoost, neural-network, and ensemble approaches reached 0.931, 0.946, 0.958, 0.949, and 0.961, respectively. This ordering was broadly consistent with earlier cybersecurity prediction studies in which ensemble and

boosting methods frequently outperformed individual decision trees because they captured nonlinear interactions while reducing instability associated with single-tree partitions. XGBoost achieved 91.8% accuracy and an F1-score of 80.4%, while the ensemble model reached 92.1% accuracy and an F1-score of 81.4%. These results were also compatible with earlier vulnerability-exploitation research showing that exploitation cannot be represented adequately by severity variables alone because temporal information, exploit evidence, product characteristics, exposure, and environmental context contribute additional predictive value. However, this study extended model comparison beyond discrimination by examining Brier scores, log-loss, and calibration error (Sadia & Maishar Mostari, 2022; Shaiful et al., 2022; Shainer et al., 2021; Stabak & Md Shahid, 2022). The ensemble model achieved a calibration error of 0.018, while the probabilistic model produced 0.019, demonstrating that strong discrimination could coexist with reliable probability estimation. Earlier machine-learning cybersecurity studies have often concentrated primarily on accuracy, precision, recall, and F1-score, with comparatively less attention to calibration. For enterprise remediation, calibration has practical importance because a predicted probability should meaningfully represent differences in expected exploitation likelihood rather than merely generate correct rank ordering. The threshold findings further supported earlier research concerning the sensitivity-specificity trade-off. Lower thresholds increased recall but expanded the remediation population and reduced precision, whereas higher thresholds improved precision and specificity at the cost of additional false negatives (Abdullah Mohammad, 2023; Adnan & Iftekhar, 2023; Bushong et al., 2021; Syed Feroz & Md Mahamud Pervaz, 2022).

Figure 12: Database risk scoring discussion summary



Therefore, the results indicated that predictive superiority should not be interpreted through a single metric or threshold. Instead, discrimination, calibration, class imbalance, false-negative consequences, and available remediation capacity collectively determined the practical usefulness of predictive cybersecurity models.

The subgroup findings demonstrated that cybersecurity risk varied across database technologies and deployment models, although exposure, privilege, criticality, patch status, and security-control strength produced greater differentiation than technology category alone. Distributed databases recorded the highest technology-specific mean composite risk score of 72.8, followed by containerized databases at 70.6 and NoSQL systems at 67.9, while DBaaS resources produced a comparatively lower mean of 61.7. Earlier cloud-security studies have reported similar heterogeneity, particularly where distributed and containerized systems increase the number of services, identities, interfaces, and dependencies involved in database access (Goniwada, 2021d). However, the findings did not indicate that a particular database technology was inherently insecure. Rather, the higher scores corresponded with architectural complexity, connectivity, exposure, and privilege characteristics surrounding those technologies. A similar interpretation applied to deployment environments. Multi-cloud systems produced a mean risk score of 73.5 and exploitation prevalence of 18.6%, compared with 58.9 and 9.8%, respectively, in private-cloud environments. Earlier studies have associated multi-cloud and hybrid environments with increased policy heterogeneity, cross-platform identity relationships, configuration inconsistency, and expanded dependency structures, which were consistent with the quantitative patterns observed here. More pronounced differences emerged when exposure and enterprise context were examined directly. Internet-facing assets recorded an exploitation prevalence of 26.5% compared with 9.8% among internally restricted resources, while administrative or highly privileged workloads reached 27.2% compared with 10.6% for non-elevated privilege. Unpatched observations similarly recorded 22.9% exploitation compared with 8.8% among patched resources (Goniwada, 2021g). These results aligned with earlier empirical cybersecurity studies identifying accessibility, privilege, patch condition, and weak security controls as major determinants of successful exploitation. Strong security controls corresponded with a mean risk score of 56.2 and exploitation prevalence of 7.5%, compared with 77.6 and 24.8% among weakly controlled assets. Asset importance also materially affected prioritization because high- and critical-importance databases generated substantially greater remediation coverage. Collectively, these findings indicated that database type and cloud deployment provided meaningful stratification, but their risk implications were strongly conditioned by the surrounding exposure, identity, control, and business context (Sun et al., 2021).

The attack-path and dependency findings provided important evidence that cloud-native database cybersecurity risk was structurally influenced by the position of vulnerabilities within interconnected enterprise architectures. Internet-facing assets exhibited an average of 8.7 feasible attack paths and a mean shortest path of 2.6 steps, whereas internally restricted assets averaged only 3.2 paths and required 4.7 steps to reach comparable critical resources (Goniwada, 2021a). Administrative or highly privileged workloads produced an even greater average of 9.3 attack paths, reached 5.6 critical assets, and required only 1.6 privilege-escalation steps. These results were consistent with earlier attack-graph, Bayesian attack-network, and dependency-analysis studies showing that vulnerability importance increases when a weakness provides multiple feasible routes, short paths, or privilege transitions toward high-value assets. Earlier research has also demonstrated that node centrality and dependency density can reveal strategically important weaknesses that remain understated by independent vulnerability scores. The present findings reinforced this argument because workloads in the highest dependency-density quartile averaged 9.6 feasible attack paths compared with only 2.4 among workloads in the lowest quartile. Mean path probability increased from 0.112 to 0.461, while the shortest path declined from 5.2 to 2.3 steps (Baresi & Garriga, 2019). These structural differences materially changed vulnerability rankings. Among 16,441 medium-severity observations, 2,318, or 14.1%, were reclassified into high or critical contextual-risk categories after exposure, privilege, critical-

asset reachability, and dependency information were incorporated. Conversely, 3,126 vulnerabilities initially classified as high or critical by CVSS received lower contextual priority because they affected isolated, strongly controlled, or lower-criticality resources. This reordering was consistent with earlier contextual-risk research arguing that vulnerability prioritization should represent the potential attack sequence rather than treating each weakness as an independent event. The findings also explained why the attack-path model achieved a substantially stronger ROC-AUC and PR-AUC than severity-only scoring. Architectural relationships supplied information about whether an attacker could practically traverse from an exposed component to sensitive database resources. Thus, attack-path analysis did not merely add another numerical factor to conventional scoring; it changed the analytical unit from isolated vulnerability characteristics toward system-level compromise opportunities (Kehrer & Blochinger, 2019). This distinction was particularly relevant in cloud-native infrastructures where microservices, APIs, containers, service accounts, and permissions created interconnected security dependencies.

The inferential findings demonstrated that the observed model and subgroup differences were supported not only by statistical significance but also by substantial effect magnitudes. Attack-path probability produced the strongest correlation with composite risk at r = .66, followed by EPSS probability at r = .61, threat activity at r = .57, identity privilege at r = .55, asset criticality at r = .52, network exposure at r = .49, and CVSS severity at r = .43. This pattern was consistent with earlier studies showing that exploitation and contextual variables explain risk more directly than technical severity when the outcome of interest concerns observed exploitation or enterprise consequence (Gleb, 2021). The relatively weak CVSS–EPSS rank association of ρ = .29 further supported previous findings that severity and exploitation probability should not be treated as interchangeable constructs. Large standardized differences were observed between exploited and non-exploited observations, publicly exposed and restricted resources, and high- versus lower-privilege workloads, with Cohen's d values of 1.05, 1.12, and 1.25, respectively (Goniwada, 2021e). These magnitudes indicated substantial practical separation rather than effects that became significant only because of the large sample. Logistic regression provided additional support because high EPSS probability increased exploitation odds by 3.42 times, high-probability attack paths by 3.17 times, active threat intelligence by 2.76 times, administrative privilege by 2.58 times, public exposure by 2.31 times, and unpatched status by 2.12 times. Strong security controls reduced the odds to 0.46, reinforcing earlier defense-in-depth research emphasizing the protective contribution of layered controls. Ranking analysis produced equally important operational differences. CVSS captured only 42.8% of verified exploited vulnerabilities in the evaluated highest-risk segment, whereas the composite model captured 88.9%. NDCG increased from 0.587 to 0.918, and Precision@10% increased from 61.6% to 86.4%. Earlier vulnerability-prioritization research has emphasized that ranking efficiency can be more operationally meaningful than global classification accuracy because security teams typically remediate only a limited fraction of available findings (Kelly et al., 2021). The present results strongly supported that position. A scoring model could therefore achieve practical value by concentrating exploited and critical vulnerabilities near the top of a constrained remediation queue, even when complete vulnerability elimination was operationally impossible.

The validation and operational findings indicated that the superiority of integrated risk assessment was not restricted to the original analytical sample. The composite model achieved a primary ROC-AUC of 0.963, which remained at 0.959 during 10-fold cross-validation, 0.957 during independent holdout testing, and 0.948 during temporal validation. Its bootstrap 95% confidence interval of 0.956–0.969 was narrow, while the holdout calibration error remained 0.020. These results were consistent with earlier methodological studies emphasizing that cybersecurity models should be evaluated using independent and temporally separated data because random development-set performance can overestimate generalizability (Haber, 2020). The relatively small decline from primary to temporal performance indicated that the integrated model retained substantial discriminatory ability when evaluated against later observations, although the temporal decrease also reflected the changing nature of vulnerability exploitation and threat activity documented in previous research. Sensitivity testing provided additional evidence of robustness because ±10% changes in component weights retained Spearman rank correlations between 0.934 and 0.982, while even ±20% perturbations maintained

correlations between 0.887 and 0.963. This stability suggested that the model's ranking performance was not dependent on a narrowly specified weighting configuration. The most operationally important finding concerned remediation workload (Goniwada, 2021b). CVSS-based prioritization identified 28,207 observations, or 58.5% of the dataset, for priority consideration, whereas the composite model at its primary threshold reduced this requirement to 12,463 observations, or 25.8%. This represented a 55.8% reduction in remediation workload while increasing high-risk coverage from 82.4% under CVSS to 94.1%. Earlier vulnerability-management studies have similarly argued that exploitation-informed prioritization can substantially reduce remediation burden, but the present results showed that additional contextual and dependency information further concentrated consequential risk. Threshold sensitivity also clarified the trade-off involved: a lower composite threshold increased high-risk coverage to 97.2% and reduced the false-negative rate to 7.2%, whereas a higher threshold reduced workload to 19.3% but decreased coverage to 85.6%. Accordingly, the findings indicated that operational risk scoring should be interpreted as a resource-allocation mechanism rather than solely as a classification exercise (Goniwada, 2021c). Validation stability, uncertainty estimates, ranking robustness, and workload reduction collectively demonstrated that integrated cybersecurity risk scoring provided a quantitatively defensible basis for prioritizing cloud-native enterprise database vulnerabilities under constrained remediation capacity.

CONCLUSION

This study concluded that cybersecurity risk assessment for enterprise database systems in cloud-native environments was substantially more effective when technical vulnerability severity was integrated with exploitation probability, threat intelligence, asset criticality, data sensitivity, network exposure, identity privilege, security controls, and architectural dependency information. The final analytical dataset contained 48,216 vulnerability–asset observations with 98.7% data completeness, providing a broad quantitative foundation across relational, NoSQL, distributed, DBaaS, and containerized database environments. Although 58.5% of observations were classified as high or critical according to CVSS severity, only 14.4% were associated with verified exploitation, demonstrating that technical severity alone did not adequately represent observed exploitation risk. This distinction was reflected in model performance, as CVSS achieved a ROC-AUC of 0.716 and PR-AUC of 0.341, whereas EPSS, threat-informed, contextual, and attack-path approaches progressively improved discrimination. The integrated composite model achieved the strongest overall performance, with 92.6% accuracy, 78.9% precision, 86.7% recall, an F1-score of 82.6%, ROC-AUC of 0.963, and PR-AUC of 0.887. Statistical analysis further demonstrated that attack-path probability, exploitation likelihood, threat activity, identity privilege, asset criticality, and network exposure were strongly associated with composite cybersecurity risk, while the multivariable model explained 68.4% of risk-score variation. High EPSS probability increased the odds of verified exploitation by 3.42 times, high-probability attack paths by 3.17 times, administrative privilege by 2.58 times, and public exposure by 2.31 times, whereas strong security controls substantially reduced exploitation likelihood. Subgroup findings additionally showed that distributed, containerized, hybrid, and multi-cloud environments exhibited elevated risk patterns, although exposure, privilege, patch status, asset importance, and security-control strength produced greater differentiation than database technology alone. Dependency-aware analysis demonstrated the importance of system-level context by elevating 14.1% of medium-severity vulnerabilities into high or critical contextual-risk categories when attack paths and critical-asset reachability were considered. The robustness of these findings was supported by a composite-model ROC-AUC of 0.959 during cross-validation, 0.957 during independent holdout testing, and 0.948 during temporal validation, with a narrow bootstrap 95% confidence interval of 0.956–0.969. Most importantly, improved predictive performance translated into operational efficiency: the composite approach reduced the priority remediation population from 28,207 observations under broad CVSS prioritization to 12,463, representing a 55.8% workload reduction while increasing high-risk coverage from 82.4% to 94.1%. Therefore, this study established that enterprise database cybersecurity risk was multidimensional and could not be represented adequately through technical severity in isolation. Exploitation-aware, contextual, predictive, and dependency-sensitive scoring provided a more discriminative, robust, and operationally efficient quantitative basis for identifying and prioritizing consequential cybersecurity risks across complex cloud-native database infrastructures.

RECOMMENDATION

Based on the quantitative findings of this study, enterprise cybersecurity programs should adopt an integrated, context-aware, and exploitation-informed risk-scoring framework for prioritizing vulnerabilities affecting cloud-native database systems rather than relying predominantly on technical severity classifications. Vulnerability-management processes should combine CVSS severity with EPSS exploitation probability, known-exploitation evidence, current threat activity, exploit availability, vulnerability age, asset criticality, data sensitivity, internet exposure, network reachability, identity privilege, patch status, security-control strength, and architectural dependency characteristics. Particular attention should be directed toward internet-facing, highly privileged, unpatched, and weakly controlled database assets because these conditions were associated with substantially greater exploitation prevalence and risk scores. Organizations should also incorporate attack-path and dependency analysis into routine database security assessments so that vulnerabilities providing short, high-probability routes to critical databases can be identified even when their technical severity is moderate. This approach is particularly important for containerized, distributed, hybrid, and multi-cloud environments where microservices, APIs, Kubernetes workloads, service accounts, permissions, and interconnected dependencies can create indirect compromise pathways that conventional vulnerability scores may not adequately represent. Predictive models such as XGBoost and ensemble learning should be considered as complementary prioritization mechanisms because they demonstrated strong discrimination, while probability calibration, independent holdout validation, temporal testing, and sensitivity analysis should remain mandatory before model outputs are incorporated into operational remediation decisions. Risk thresholds should additionally be aligned with organizational risk tolerance and remediation capacity rather than applied uniformly. The composite model's primary threshold reduced the priority remediation population from 28,207 to 12,463 observations, representing a 55.8% workload reduction while retaining 94.1% high-risk coverage; however, organizations with lower tolerance for missed critical vulnerabilities should consider lower thresholds that increase coverage even when remediation workload rises. Security teams should continuously reassess high-value database assets whenever exposure, privilege, dependencies, threat intelligence, or security controls change because contextual risk can vary without any change in the underlying CVSS score. Strong identity and access management, least-privilege enforcement, network segmentation, multifactor authentication, secrets protection, encryption, timely patching, and reduction of unnecessary public endpoints should accompany quantitative scoring because strong security controls were associated with materially lower exploitation likelihood. Finally, organizations should evaluate risk-scoring effectiveness through precision, recall, F1-score, ROC-AUC, PR-AUC, calibration measures, false-negative rates, Precision@K, Recall@K, ranking quality, remediation coverage, and workload reduction rather than depending on accuracy or severity distributions alone. Such an integrated governance approach would support more defensible allocation of limited cybersecurity resources toward vulnerabilities presenting the greatest combination of exploitation likelihood, architectural accessibility, and enterprise consequence.

LIMITATIONS

This study contained several limitations that should be considered when interpreting the quantitative assessment of cybersecurity risk-scoring models for enterprise database systems in cloud-native environments. First, the analytical dataset contained 48,216 retained vulnerability–asset observations with 98.7% data completeness, but the observations could not represent every database technology, cloud provider, organizational architecture, threat environment, or security configuration encountered across global enterprises. Relational databases constituted the largest technology group, while distributed and other specialized cloud-connected databases represented smaller proportions, potentially producing unequal subgroup representation. Second, verified exploitation was observed for 14.4% of the dataset, creating meaningful class imbalance between exploited and non-exploited observations. Although PR-AUC, precision, recall, F1-score, stratified validation, and class-sensitive evaluation were applied to reduce misleading interpretations, rare exploitation patterns could still have been underrepresented. Third, exploitation labels depended on available vulnerability, threat-intelligence, known-exploitation, and organizational security information; therefore, exploitation events that remained undetected, unreported, incorrectly attributed, or absent from available records

could have introduced outcome misclassification. Fourth, contextual variables such as asset criticality, data sensitivity, security-control strength, network exposure, and privilege were operationalized through structured classifications, and some degree of measurement variation could have remained across heterogeneous enterprise environments. Fifth, attack-path analysis depended on the completeness of network, identity, service, permission, and dependency relationships. Missing or dynamically changing cloud connections could have affected estimated path counts, shortest paths, privilege transitions, reachable critical assets, and attack probabilities. Sixth, the composite risk model incorporated multiple weighted dimensions, and although sensitivity analysis demonstrated strong ranking stability under ±10% and ±20% weight modifications, alternative weighting structures could still produce different prioritization outcomes in organizations with different operational objectives and risk tolerances. Seventh, machine-learning models, particularly XGBoost, neural-network, and ensemble approaches, could have captured dataset-specific relationships that were not fully transferable to substantially different enterprise environments. Cross-validation, independent holdout testing, bootstrap confidence intervals, and temporal validation reduced this concern but did not eliminate external-validity limitations.

REFERENCES

[1]. Abdullah Mohammad, S. (2023). $CUGAO_2$ and $CUINS_2$ Quantum Dot Sensitization for Enhanced Photovoltaic Efficiency in Wearable Fiber-Shaped Solar Cells. *American Journal of Advanced Technology and Engineering Solutions*, *3*(04), 209-261. https://doi.org/10.63125/wysdvy71

[2]. Aditya, K., Grzonkowski, S., & Le-Khac, N.-A. (2018). Riskwriter: Predicting cyber risk of an enterprise. International conference on information systems security,

[3]. Adnan, S., & Iftekhar, A. (2023). Predicting Cost Variance in Digital Transformation Projects Through Ensemble Learning and Project Performance Analytics. *International Journal of Scientific Interdisciplinary Research*, *13*(10), 01–36. https://doi.org/10.63125/j18hxs51

[4]. Ahmed, H. C. H. R., & Nuzhat Sadia, P. (2023). A Quantitative Assessment of CRM Analytics and KPI Dashboard Design on Enterprise Revenue Conversion in Emerging Market Logistics Platforms. *American Journal of Data Science and Analytics*, *4*(09), 50-95. https://doi.org/10.63125/315bge15

[5]. Algarni, A. M., Thayananthan, V., & Malaiya, Y. K. (2021). Quantitative assessment of cybersecurity risks for mitigating data breaches in business systems. *Applied Sciences*, *11*(8), 3678.

[6]. Ali, S. (2021). Cybersecurity management for distributed control system: systematic approach. *Journal of Ambient Intelligence and Humanized Computing*, *12*(11), 10091-10103.

[7]. Allodi, L., Cremonini, M., Massacci, F., & Shim, W. (2020). Measuring the accuracy of software vulnerability assessments: experiments with students and professionals. *Empirical Software Engineering*, *25*(2), 1063-1094.

[8]. Andreolini, M., Colacino, V. G., Colajanni, M., & Marchetti, M. (2020). A framework for the evaluation of trainee performance in cyber range exercises. *Mobile Networks and Applications*, *25*(1), 236-247.

[9]. Armenia, S., Angelini, M., Nonino, F., Palombi, G., & Schlitzer, M. F. (2021). A dynamic simulation approach to support the evaluation of cyber risks and security investments in SMEs. *Decision Support Systems*, *147*, 113580.

[10]. Asheke, M. (2023). Machine Learning–Based Predictive Quality Modeling for Defect Detection and Process Optimization in Advanced Manufacturing Systems. *Review of Applied Science and Technology*, *13*(12), 01–49. https://doi.org/10.63125/thy73b16

[11]. Ashik, M., & Shabbir, A. (2022). Digital Twin-Enabled Predictive Maintenance for Rotating and Static Equipment: Bridging FEA-Based Failure Analysis with Real-Time Asset Health Monitoring in Energy and Power Facilities. *International Journal of Scientific Interdisciplinary Research*, *12*(05), 01–37. https://doi.org/10.63125/0qaysz49

[12]. Baresi, L., & Garriga, M. (2019). Microservices: The evolution and extinction of web services? In *Microservices: Science and Engineering* (pp. 3-28). Springer.

[13]. Bauer, M. S., Chadha, R., & Viswanathan, M. (2020). Modelchecking safety properties in randomized security protocols. In *Logic, Language, and Security: Essays Dedicated to Andre Scedrov on the Occasion of His 65th Birthday* (pp. 167-183). Springer.

[14]. Baybutt, P. (2017). Issues for security risk assessment in the process industries. *Journal of Loss Prevention in the Process Industries*, *49*, 509-518.

[15]. Bharathi, S. V. (2017). Prioritizing and ranking the big data information security risk spectrum. *Global Journal of Flexible Systems Management*, *18*(3), 183-201.

[16]. Bird, D. (2020). Derivation of a conceptual framework to assess and mitigate identified customer cybersecurity risks by utilizing the public cloud. Fourth International Congress on Information and Communication Technology: ICICT 2019, London, Volume 2,

[17]. Bloomfield, R. E., Popov, P., Salako, K., Stankovic, V., & Wright, D. (2017). Preliminary interdependency analysis: An approach to support critical-infrastructure risk-assessment. *Reliability engineering & system safety*, *167*, 198-217.

[18]. Borky, J. M., & Bradley, T. H. (2018). Protecting information with cybersecurity. In *Effective model-based systems engineering* (pp. 345-404). Springer.

[19]. Broadstock, D. C., Cheng, L. T., & Poon, J. S. (2021). Fintech unicorns. In *The Palgrave Handbook of FinTech and Blockchain* (pp. 109-170). Springer.

[20]. Bushong, V., Abdelfattah, A. S., Maruf, A. A., Das, D., Lehman, A., Jaroszewski, E., Coffey, M., Cerny, T., Frajtak, K., & Tisnovsky, P. (2021). On microservice analysis and architecture evolution: A systematic mapping study. *Applied Sciences*, *11*(17), 7856.

[21]. Cantelmi, R., Di Gravio, G., & Patriarca, R. (2021). Reviewing qualitative research approaches in the context of critical infrastructure resilience. *Environment Systems and Decisions*, *41*(3), 341-376.

[22]. Cao, C., Yuan, L.-P., Singhal, A., Liu, P., Sun, X., & Zhu, S. (2018). Assessing attack impact on business processes by interconnecting attack graphs and entity dependency graphs. IFIP Annual Conference on Data and Applications Security and Privacy,

[23]. Chatterjee, S., & Thekdi, S. (2020). An iterative learning and inference approach to managing dynamic cyber vulnerabilities of complex systems. *Reliability engineering & system safety*, *193*, 106664.

[24]. Chia, K. S., Jam, M. N. H., Gan, Z., & Ismail, N. (2020). Pre-dispersive near-infrared light sensing in non-destructively classifying the brix of intact pineapples. *Journal of Food Science and Technology*, *57*(12), 4533-4540.

[25]. Christen, M., Gordijn, B., Weber, K., Van de Poel, I., & Yaghmaei, E. (2017). A review of value-conflicts in cybersecurity: an assessment based on quantitative and qualitative literature analysis. *The ORBIT Journal*, *1*(1), 1-19.

[26]. Cococcioni, M., Fiaschi, L., & Lermusiaux, P. F. (2021). Game theory for unmanned vehicle path planning in the marine domain: State of the art and new possibilities. *Journal of Marine Science and Engineering*, *9*(11), 1175.

[27]. Cui, J., Wang, L., Zhao, X., & Zhang, H. (2020). Towards predictive analysis of android vulnerability using statistical codes and machine learning for IoT applications. *Computer Communications*, *155*, 125-131.

[28]. De Silva, K., Lee, W. K., Forbes, A., Demmer, R. T., Barton, C., & Enticott, J. (2020). Use and performance of machine learning models for type 2 diabetes prediction in community settings: A systematic review and meta-analysis. *International journal of medical informatics*, *143*, 104268.

[29]. Derbyshire, R., Green, B., & Hutchison, D. (2021). "Talking a different Language": Anticipating adversary attack cost for cyber risk assessment. *Computers & Security*, *103*, 102163.

[30]. Diez-Olivan, A., Pagan, J. A., Sanz, R., & Sierra, B. (2017). Data-driven prognostics using a combination of constrained K-means clustering, fuzzy modeling and LOF-based score. *Neurocomputing*, *241*, 97-107.

[31]. Dobaj, J., Schmittner, C., Krisper, M., & Macher, G. (2019). Towards integrated quantitative security and safety risk assessment. International Conference on Computer Safety, Reliability, and Security,

[32]. Fenlon, C., O'Grady, L., Doherty, M. L., & Dunnion, J. (2018). A discussion of calibration techniques for evaluating binary and categorical predictive models. *Preventive veterinary medicine*, *149*, 107-114.

[33]. Fielder, A., König, S., Panaousis, E., Schauer, S., & Rass, S. (2018). Risk assessment uncertainties in cybersecurity investments. *Games*, *9*(2), 34.

[34]. Figueira, P. T., Bravo, C. L., & López, J. L. R. (2020). Improving information security risk analysis by including threat-occurrence predictive models. *Computers & Security*, *88*, 101609.

[35]. Gadyatskaya, O., Jhawar, R., Mauw, S., Trujillo-Rasua, R., & Willemse, T. A. (2017). Refinement-aware generation of attack trees. International Workshop on Security and Trust Management,

[36]. Garriga, M. (2017). Towards a taxonomy of microservices architectures. International conference on software engineering and formal methods,

[37]. Gawinski, L., Engelseth, P., & Kozlowski, R. (2021). Application of modern clinical risk scores in the global assessment of risks related to the diagnosis and treatment of acute coronary syndromes in everyday medical practice. *International Journal of Environmental Research and Public Health*, *18*(17), 9103.

[38]. Gazi Yeasin, A. (2023). Impact of Accessibility-First Design Systems on WCAG 2.1 Compliance and Inclusive User Experience in Enterprise Software Platforms: A Mixed-Methods Evaluation. *Review of Applied Science and Technology*, *13*(07), 01–49. https://doi.org/10.63125/nxh4gp19

[39]. George, G., & Thampi, S. M. (2019). Vulnerability-based risk assessment and mitigation strategies for edge devices in the Internet of Things. *Pervasive and Mobile Computing*, *59*, 101068.

[40]. Giuca, O., Popescu, T. M., Popescu, A. M., Prostean, G., & Popescu, D. E. (2018). A survey of cybersecurity risk management frameworks. International Workshop Soft Computing Applications,

[41]. Gleb, T. (2021). Procure Software Delivery Environment. In *Systematic Cloud Migration: A Hands-On Guide to Architecture, Design, and Technical Implementation* (pp. 143-160). Springer.

[42]. Gonda, T., Puzis, R., & Shapira, B. (2017). Scalable attack path finding for increased security. International Conference on Cyber Security Cryptography and Machine Learning,

[43]. Goniwada, S. R. Cloud Native Architecture and Design.

[44]. Goniwada, S. R. (2021a). Cloud native architecture and design patterns. In *Cloud Native Architecture and Design: A Handbook for Modern Day Architecture and Design with Enterprise-Grade Examples* (pp. 127-187). Springer.

[45]. Goniwada, S. R. (2021b). Cloud Native Architecture Principles. In *Cloud Native Architecture and Design: A Handbook for Modern Day Architecture and Design with Enterprise-Grade Examples* (pp. 55-125). Springer.

[46]. Goniwada, S. R. (2021c). Containerization and virtualization. In *Cloud Native Architecture and Design: A Handbook for Modern Day Architecture and Design with Enterprise-Grade Examples* (pp. 573-617). Springer.

[47]. Goniwada, S. R. (2021d). Designing for "-ilities". In *Cloud Native Architecture and Design: A Handbook for Modern Day Architecture and Design with Enterprise-Grade Examples* (pp. 371-409). Springer.

[48]. Goniwada, S. R. (2021e). Enterprise cloud native automation. In *Cloud Native Architecture and Design: A Handbook for Modern Day Architecture and Design with Enterprise-Grade Examples* (pp. 523-553). Springer.

[49]. Goniwada, S. R. (2021f). Introduction to Cloud Native Architecture. In *Cloud Native Architecture and Design: A Handbook for Modern Day Architecture and Design with Enterprise-Grade Examples* (pp. 3-26). Springer.

[50]. Goniwada, S. R. (2021g). Observability. In *Cloud native architecture and design: a handbook for modern day architecture and design with enterprise-grade examples* (pp. 661-676). Springer.
[51]. Gonzalez-Granadillo, G., Menesidou, S. A., Papamartzivanos, D., Romeu, R., Navarro-Llobet, D., Okoh, C., Nifakos, S., Xenakis, C., & Panaousis, E. (2021). Automated cyber and privacy risk management toolkit. *Sensors*, *21*(16), 5493.
[52]. Granata, D., Rak, M., & Salzillo, G. (2021). Risk analysis automation process in it security for cloud applications. International Conference on Cloud Computing and Services Science,
[53]. Grünewald, E. (2021). Cloud native privacy engineering through DevPrivOps. In *IFIP International Summer School on Privacy and Identity Management* (pp. 122-141). Springer.
[54]. Haber, M. J. (2020). The Cloud. In *Privileged Attack Vectors: Building Effective Cyber-Defense Strategies to Protect Organizations* (pp. 215-226). Springer.
[55]. Hacker, P., & Passoth, J.-H. (2020). Varieties of AI explanations under the law. From the GDPR to the AIA, and beyond. International workshop on extending explainable AI beyond deep models and classifiers,
[56]. Haley, M. R. (2017). K-fold cross validation performance comparisons of six naive portfolio selection rules: how naive can you be and still have successful out-of-sample portfolio performance? *Annals of Finance*, *13*(3), 341-353.
[57]. Han, C.-H., & Han, C. (2021). Semi-quantitative cybersecurity risk assessment by blockade and defense level analysis. *Process Safety and Environmental Protection*, *155*, 306-316.
[58]. Harirchian, E., Kumari, V., Jadhav, K., Rasulzade, S., Lahmer, T., & Raj Das, R. (2021). A synthesized study based on machine learning approaches for rapid classifying earthquake damage grades to RC buildings. *Applied Sciences*, *11*(16), 7540.
[59]. Hermanowski, D., & Piotrowski, R. (2021). Network risk assessment based on attack graphs. International Conference on Dependability and Complex Systems,
[60]. Humayun, M., Niazi, M., Jhanjhi, N. Z., Alshayeb, M., & Mahmood, S. (2020). Cyber security threats and vulnerabilities: a systematic mapping study. *Arabian Journal for Science and Engineering*, *45*(4), 3171-3189.
[61]. Jahanara, A., & Mst Kaniz, F. (2023). Demand Variability and Capacity Planning in Home and Community-Based Services: A Decision-Analytic Framework for Health Services Managers. *American Journal of Data Science and Analytics*, *4*(09), 01-49. https://doi.org/10.63125/3y62vr16
[62]. Jia, Y., Qi, Y., Shang, H., Jiang, R., & Li, A. (2018). A practical approach to constructing a knowledge graph for cybersecurity. *Engineering*, *4*(1), 53-60.
[63]. Jouini, M., Ben Arfa Rabai, L., & Khedri, R. (2021). A quantitative assessment of security risks based on a multifaceted classification approach: M. Jouini et al. *International Journal of Information Security*, *20*(4), 493-510.
[64]. Kandasamy, K., Srinivas, S., Achuthan, K., & Rangan, V. P. (2020). IoT cyber risk: A holistic analysis of cyber risk assessment frameworks, risk vectors, and risk ranking process. *EURASIP Journal on Information Security*, *2020*(1), 8.
[65]. Kara, M. E., Fırat, S. Ü. O., & Ghadge, A. (2020). A data mining-based framework for supply chain risk management. *Computers & Industrial Engineering*, *139*, 105570.
[66]. 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
[67]. Kehrer, S., & Blochinger, W. (2019). Migrating parallel applications to the cloud: assessing cloud readiness based on parallel design decisions: S. Kehrer, W. Blochinger. *SICS Software-Intensive Cyber-Physical Systems*, *34*(2), 73-84.
[68]. Kelly, D., Glavin, F. G., & Barrett, E. (2021). Denial of wallet—defining a looming threat to serverless computing. *Journal of Information Security and Applications*, *60*, 102843.
[69]. Khan, F., Kim, J. H., Mathiassen, L., & Moore, R. (2021). Data breach management: an integrated risk model. *Information & Management*, *58*(1), 103392.
[70]. Khan, M. A., Farooq, F., Javed, M. F., Zafar, A., Ostrowski, K. A., Aslam, F., Malazdrewicz, S., & Maślak, M. (2021). Simulation of depth of wear of eco-friendly concrete using machine learning based computational approaches. *Materials*, *15*(1), 58.
[71]. Kiani, J., Camp, C., & Pezeshk, S. (2019). On the application of machine learning techniques to derive seismic fragility curves. *Computers & structures*, *218*, 108-122.
[72]. Kim, Y. S., Choi, M. K., Han, S. M., Lee, C., & Seong, P. H. (2020). Development of a method for quantifying relative importance of NPP cyber attack probability variables based on factor analysis and AHP. *Annals of Nuclear Energy*, *149*, 107790.
[73]. Kure, H. I., Islam, S., & Razzaque, M. A. (2018). An integrated cyber security risk management approach for a cyber-physical system. *Applied Sciences*, *8*(6), 898.
[74]. Kuzminykh, I., Ghita, B., Sokolov, V., & Bakhshi, T. (2021). Information security risk assessment. *Encyclopedia*, *1*(3), 602-617.
[75]. Lazăr, D. C., Avram, M. F., Faur, A. C., Goldiş, A., Romoşan, I., Tăban, S., & Cornianu, M. (2020). The impact of artificial intelligence in the endoscopic assessment of premalignant and malignant esophageal lesions: present and future. *Medicina*, *56*(7), 364.
[76]. Lee, G., Kim, W., Oh, H., Youn, B. D., & Kim, N. H. (2019). Review of statistical model calibration and validation—from the perspective of uncertainty structures: Guesuk Lee et al. *Structural and Multidisciplinary Optimization*, *60*(4), 1619-1644.
[77]. Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, *7*(1), 29.

[78]. Li, W., Li, B., Fang, R., You, P., Zou, Y., Xu, Z., & Guo, S. (2021). Risk evaluation of electric power grid enterprise related to electricity transmission and distribution tariff regulation employing a hybrid MCDM model. *Mathematics*, *9*(9), 989.

[79]. Lokanan, M., & Liu, S. (2021). Predicting fraud victimization using classical machine learning. *Entropy*, *23*(3), 300.

[80]. 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

[81]. Maleki, F., Muthukrishnan, N., Ovens, K., Reinhold, C., & Forghani, R. (2020). Machine learning algorithm validation: from essentials to advanced applications and implications for regulatory certification and deployment. *Neuroimaging Clinics of North America*, *30*(4), 433-445.

[82]. Mamonov, S., & Triantoro, T. M. (2018). The strategic value of data resources in emergent industries. *International Journal of Information Management*, *39*, 146-155.

[83]. 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

[84]. Mashiat, N., & Md Anisur Rahman, M. (2022). Machine Learning Approaches to Credit Risk Scoring and Default Prediction in Small-Business Lending: Empirical Evidence from Lending Portfolio Data. *American Journal of Data Science and Analytics*, *3*(06), 137-182. https://doi.org/10.63125/j256rs96

[85]. Mateo Tudela, F., Bermejo Higuera, J.-R., Bermejo Higuera, J., Sicilia Montalvo, J.-A., & Argyros, M. I. (2020). On combining static, dynamic and interactive analysis security testing tools to improve owasp top ten security vulnerability detection in web applications. *Applied Sciences*, *10*(24), 9119.

[86]. Matthews, I., Soudjani, S., & van Moorsel, A. (2021). Stochastic simulation techniques for inference and sensitivity analysis of Bayesian attack graphs. International Conference on Science of Cyber Security,

[87]. Md, A., & Gazi Yeasin, A. (2022). Machine Learning Enhanced Spatial-Data Workflow Redesign: A Quantitative Evaluation of Task Success and Operational Cost Reduction in Enterprise LiDAR Platforms. *American Journal of Data Science and Analytics*, *3*(06), 80-136. https://doi.org/10.63125/0den7c02

[88]. Md Anisur Rahman, M. (2023). Development of AI-Driven Business Intelligence Frameworks for Financial Risk and Liquidity Monitoring in Real Time. *International Journal of Scientific Interdisciplinary Research*, *4*(2), 152–203. https://doi.org/10.63125/jm55ca64

[89]. 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

[90]. Md Asif Ali Sheak, A., & Tahmedur, R. (2023). A Data-Driven Framework for Engineering Project Risk Assessment Under Schedule and Cost Uncertainty. *International Journal of Scientific Interdisciplinary Research*, *4*(2), 204–251. https://doi.org/10.63125/kq4zzh16

[91]. Md Kaium, H. (2023). A GIS-Based Multi-Criteria Decision-Support and Land-Suitability Optimization Framework for Climate-Resilient Urban and Regional Planning. *American Journal of Data Science and Analytics*, *4*(06), 43–103. https://doi.org/10.63125/wvq0sx87

[92]. Md Mahamud Pervaz, P. (2023). Integration of MES, SAP, And PI Historian Platforms for Real-Time Industrial Asset Performance Management. *Review of Applied Science and Technology*, *2*(04), 393–445. https://doi.org/10.63125/9ebtas97

[93]. Meland, P. H., Tokas, S., Erdogan, G., Bernsmed, K., & Omerovic, A. (2021). A systematic mapping study on cyber security indicator data. *Electronics*, *10*(9), 1092.

[94]. Miaoui, Y., & Boudriga, N. (2019). Enterprise security investment through time when facing different types of vulnerabilities. *Information Systems Frontiers*, *21*(2), 261-300.

[95]. Mohammad Badrul, A., & Md. Mominul, H. (2023). Machine Learning–Based Assessment of Environmental and Public Health Risks in Sewerage Sludge Management Systems. *American Journal of Advanced Technology and Engineering Solutions*, *3*(02), 33-77. https://doi.org/10.63125/y3yfxa92

[96]. Molla Al Rakib, H., & Mohammad Towhidul, I. (2023). Cyber-Resilient Marketing Intelligence Systems for Fraud Detection and Data Integrity in Digital Campaigns. *American Journal of Scholarly Research and Innovation*, *2*(02), 485–518. https://doi.org/10.63125/b53pgr11

[97]. Mst Kaniz, F., & Jahanara, A. (2022). Lean Six Sigma Implementation in Home Health Administrative Operations: A DMAIC Case Study of Managerial Workflow Redesign and Error Reduction. *International Journal of Scientific Interdisciplinary Research*, *12*(04), 01–60. https://doi.org/10.63125/xchca204

[98]. Mukut Kanti, B. (2023). Importance of the Sustainable growth in the Testing, Inspection and Certification (TIC) industries through AI tools and techniques. *Review of Applied Science and Technology*, *13*(07), 50–75. https://doi.org/10.63125/0aagev30

[99]. Myszkowski, N., & Storme, M. (2017). Measuring "good taste" with the visual aesthetic sensitivity test-revised (VAST-R). *Personality and Individual Differences*, *117*, 91-100.

[100]. Naseer, A., Naseer, H., Ahmad, A., Maynard, S. B., & Siddiqui, A. M. (2021). Real-time analytics, incident response process agility and enterprise cybersecurity performance: A contingent resource-based analysis. *International Journal of Information Management*, *59*, 102334.

[101]. Nishat, J. (2023). Communicable Disease Prevention in a Post-Pandemic Era: A Global Review of Public Health Education, Surveillance, and Community-Level Interventions. *International Journal of Scientific Interdisciplinary Research*, *4*(4), 630–678. https://doi.org/10.63125/mnqefj36

[102]. Nuzhat Sadia, P. (2023). A Multi-Criteria Decision-Support Framework For ERP-Based Business Process Optimization in Mid-Sized Manufacturing Enterprises. *American Journal of Scholarly Research and Innovation*, *2*(02), 519–573. https://doi.org/10.63125/j58k5d55

[103]. Oka, A., Ishimura, N., & Ishihara, S. (2021). A new dawn for the use of artificial intelligence in gastroenterology, hepatology and pancreatology. *Diagnostics*, *11*(9), 1719.

[104]. Papamartzivanos, D., Menesidou, S. A., Gouvas, P., & Giannetsos, T. (2021). A perfect match: Converging and automating privacy and security impact assessment on-the-fly. *Future Internet*, *13*(2), 30.

[105]. Park, J., Choo, Y., & Lee, J. (2019). A Hybrid Vulnerability Analysis Tool Using a Risk Evaluation Technique: J.-P. Park et al. *Wireless Personal Communications*, *105*(2), 443-459.

[106]. Park, J. W., & Lee, S. J. (2020). A quantitative assessment framework for cyber-attack scenarios on nuclear power plants using relative difficulty and consequence. *Annals of Nuclear Energy*, *142*, 107432.

[107]. Peralta, R., Garbelli, M., Bellocchio, F., Ponce, P., Stuard, S., Lodigiani, M., Fazendeiro Matos, J., Ribeiro, R., Nikam, M., & Botler, M. (2021). Development and validation of a machine learning model predicting arteriovenous fistula failure in a large network of dialysis clinics. *International Journal of Environmental Research and Public Health*, *18*(23), 12355.

[108]. Qie, Z., & Rong, L. (2017). An integrated relative risk assessment model for urban disaster loss in view of disaster system theory. *Natural Hazards*, *88*(1), 165-190.

[109]. Radanliev, P., De Roure, D. C., Nicolescu, R., Huth, M., Montalvo, R. M., Cannady, S., & Burnap, P. (2018). Future developments in cyber risk assessment for the internet of things. *Computers in industry*, *102*, 14-22.

[110]. Rahman, H. T., Robinson, B. E., Ford, J. D., & Hickey, G. M. (2018). How do capital asset interactions affect livelihood sensitivity to climatic stresses? Insights from the northeastern floodplains of Bangladesh. *Ecological Economics*, *150*, 165-176.

[111]. Rea-Guaman, A. M., Mejía, J., San Feliu, T., & Calvo-Manzano, J. A. (2020). AVARCIBER: a framework for assessing cybersecurity risks. *Cluster Computing*, *23*(3), 1827-1843.

[112]. Risha, A., & Mst Shurovi, A. (2023). Cloud-Based Accounting Systems and Cybersecurity Frameworks for Financial Data Protection in Emerging Economies: A Meta-Analysis. *International Journal of Business and Economics Insights*, *3*(4), 26–64. https://doi.org/10.63125/zb0h3n85

[113]. Ruan, K. (2017). Introducing cybernomics: A unifying economic framework for measuring cyber risk. *Computers & Security*, *65*, 77-89.

[114]. 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

[115]. Sagar, R., Jhaveri, R., & Borrego, C. (2020). Applications in security and evasions in machine learning: a survey. *Electronics*, *9*(1), 97.

[116]. Sales, T. P., Baião, F., Guizzardi, G., Almeida, J. P. A., Guarino, N., & Mylopoulos, J. (2018). The common ontology of value and risk. International conference on conceptual modeling,

[117]. Saxena, N., Hayes, E., Bertino, E., Ojo, P., Choo, K.-K. R., & Burnap, P. (2020). Impact and key challenges of insider threats on organizations and critical businesses. *Electronics*, *9*(9), 1460.

[118]. 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

[119]. Shainer, G., Graham, R., Newburn, C. J., Hernandez, O., Bloch, G., Gibbs, T., & Wells, J. C. (2021). Nvidia's cloud native supercomputing. Smoky Mountains Computational Sciences and Engineering Conference,

[120]. Sheehan, B., Murphy, F., Mullins, M., & Ryan, C. (2019). Connected and autonomous vehicles: A cyber-risk classification framework. *Transportation research part A: policy and practice*, *124*, 523-536.

[121]. Stabak, D., & Md Shahid, N. (2022). Empirical Assessment of Uncertainty-Aware Real-Time Decision-Making for Reliable and Resilient Cyber-Physical Systems. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 432–477. https://doi.org/10.63125/q3r4kk16

[122]. Stabak, D., & Md Siam, T. (2023). Real-Time Edge Computing for Cyber-Physical Systems: A Meta-Analysis of Energy Efficiency and Fault Tolerance. *American Journal of Scholarly Research and Innovation*, *13*(06), 01–44. https://doi.org/10.63125/g9c9cp41

[123]. Sun, R., Wang, Q., & Guo, L. (2021). Research towards key issues of API security. China Cyber Security Annual Conference,

[124]. Syed Feroz, A. (2023). Quality-Management Systems for Resilient Energy Infrastructure Fabrication. *International Journal of Scientific Interdisciplinary Research*, *13*(10), 37–91. https://doi.org/10.63125/mbhhrm88

[125]. Syed Feroz, A., & Md Mahamud Pervaz, P. (2022). Root-Cause Analysis and FMEA-Driven Reliability Improvement in Gas Production Facilities: A Multi-Year Empirical Study. *International Journal of Scientific Interdisciplinary Research*, *12*(10), 01–53. https://doi.org/10.63125/70msp878

[126]. Tahmedur, R., & Rifat, C. (2023). Critical Path Analysis and Resource-Leveling Optimization for Multi-Project Engineering Portfolios. *American Journal of Interdisciplinary Studies*, *13*(12), 01-47. https://doi.org/10.63125/8eqnea46

[127]. Tal, E. (2017). Calibration: Modelling the measurement process. *Studies in History and Philosophy of Science Part A*, *65*, 33-45.

[128]. Tehrany, M. S., Jones, S., Shabani, F., Martínez-Álvarez, F., & Tien Bui, D. (2019). A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data: MS Tehrany et al. *Theoretical and Applied Climatology*, *137*(1), 637-653.
[129]. Torkura, K. A., Sukmana, M. I., Cheng, F., & Meinel, C. (2018). Cavas: Neutralizing application and container security vulnerabilities in the cloud native era. International Conference on Security and Privacy in Communication Systems,
[130]. Torkura, K. A., Sukmana, M. I., Cheng, F., & Meinel, C. (2021). Continuous auditing and threat detection in multi-cloud infrastructure. *Computers & Security*, *102*, 102124.
[131]. Tsamardinos, I., Greasidou, E., & Borboudakis, G. (2018). Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. *Machine learning*, *107*(12), 1895-1922.
[132]. Uddin, M. H., Ali, M. H., & Hassan, M. K. (2020). Cybersecurity hazards and financial system vulnerability: a synthesis of literature: Md. H. Uddin et al. *Risk Management*, *22*(4), 239-309.
[133]. Ulven, J. B., & Wangen, G. (2021). A systematic review of cybersecurity risks in higher education. *Future Internet*, *13*(2), 39.
[134]. Vettoretti, M., & Di Camillo, B. (2021). A variable ranking method for machine learning models with correlated features: In-silico validation and application for diabetes prediction. *Applied Sciences*, *11*(16), 7740.
[135]. Wagner, I., & Boiten, E. (2018). Privacy risk assessment: from art to science, by metrics. International Workshop on Data Privacy Management,
[136]. Wang, L., & Zhao, J. (2020). *Strategic blueprint for enterprise analytics*. Springer.
[137]. Wangen, G. (2019). Quantifying and analyzing information security risk from incident data. International Workshop on Graphical Models for Security,
[138]. Willumsen, P., Oehmen, J., Stingl, V., & Geraldi, J. (2019). Value creation through project risk management. *International Journal of Project Management*, *37*(5), 731-749.
[139]. Xiong, W. (2021). *Enhancing IT Systems Cyber Resilience through Threat Modeling: Cyber Security Analysis of Enterprise Systems and Connected Vehicles* KTH Royal Institute of Technology].
[140]. Yang, H., Zhang, S., Liu, R., Krall, A., Wang, Y., Ventura, M., & Deflitch, C. (2020). Epidemic informatics and control: A review from system informatics to epidemic response and risk management in public health. INFORMS International Conference on Service Science,
[141]. Yin, J., Tang, M., Cao, J., & Wang, H. (2020). Apply transfer learning to cybersecurity: Predicting exploitability of vulnerabilities by description. *Knowledge-Based Systems*, *210*, 106529.
[142]. Yoo, Y., & Park, H.-S. (2021). Qualitative risk assessment of cybersecurity and development of vulnerability enhancement plans in consideration of digitalized ship. *Journal of Marine Science and Engineering*, *9*(6), 565.