ISSN 2723-6471

A Hybrid Intelligent Cybersecurity Assessment System For Electronic Document Management Systems

Assylzhan Svanov[1], Assel Omarbekova[2,*], Gulmira Bekmanova[3], Alibek Barlybayev[4], Bibigul Razakhova[5], Lena Zhetkenbay[6], Magripa Saukhanova[7], Aizhan Nazyrova[8,*], Zhanar Lamasheva[9,*]

[1,2,3,4,5,6,7,8,9]*L.N. Gumilyov Eurasian National University, 2 Satpayev Str., 010008 Astana, Kazakhstan*

(Received: February 20, 2026; Revised: April 25, 2026; Accepted: July 20, 2026; Available online: August 18, 2026)

Abstract

Electronic document management systems concentrate confidential and commercially sensitive information behind a single perimeter, making them a high-priority cyberattack target, while existing assessment methods remain largely static, checklist-based, or single-technique, poorly capturing configuration dynamics or prioritizing measures by expected risk reduction. The objective is to develop and validate an intelligent system for the quantitative, explainable cybersecurity assessment of such systems. The novelty and contribution are a hybrid architecture jointly integrating Mamdani fuzzy inference, a stacking ensemble of machine-learning classifiers (random forest, gradient boosting, and a multilayer perceptron), and a Bayesian threat network with an attack graph, unified by a shared domain ontology of document-management assets and threats weighted by confidentiality, integrity, and availability. These heterogeneous estimates are aggregated into a composite cybersecurity assessment index via a fuzzy analytic hierarchy process with adaptive re-weighting from confirmed incidents, while a two-level explainability layer traces each score to its features, fired rules, and probable attack paths. The system was evaluated on 10,239 labelled configuration states from an operational deployment using cross-validation, comparison against seven baseline models, and an ablation study. It achieved the best results among all compared approaches, with an F1-score of 0.946, a Matthews correlation coefficient of 0.927, and an area under the ROC curve of 0.972 – a gain of 2.4 percentage points over the strongest baseline and 10.4 over logistic regression – and the ablation study confirmed that the fuzzy and Bayesian components contribute complementary gains. These findings show that combining data-driven learning with expert-interpretable reasoning yields a more accurate and stable assessment than any single paradigm, and indicate that the framework can support continuous, evidence-based cybersecurity monitoring in document-centric organizations, with future work on adversarial robustness and multi-organization validation.

Keywords: Cybersecurity, Electronic Document Management System, Risk Assessment, Fuzzy Logic, Machine Learning, Fuzzy AHP, Bayesian Network, Explainable AI.

1. Introduction

Electronic document management has become a backbone of digital governance. Electronic document management systems (EDMS) provide document storage and routing, the legally significant recording of decisions, access-rights management, and integration with external services. The concentration of confidential, personal, and commercially sensitive information within a single perimeter turns EDMS into a high-priority target for attackers and an object of heightened regulatory requirements; crucially, the value resides not so much in the infrastructure as in the content, versions, and access rights of documents [1], [2].

EDMS-specific threats include credential compromise and privilege escalation, unauthorized access to the repository, violation of version integrity, leakage through integration gateways, and availability attacks. Static, predominantly qualitative assessment based on checklists [3], [4], [5] poorly reflects the dynamics of the configuration state and does not allow measures to be prioritized by expected risk reduction. Existing approaches apply a single class of methods in isolation: fuzzy systems do not learn from data [6], [7], machine-learning models are opaque [8], [9], [10], and Bayesian networks require prior distributions [11], [12]. The absence of a unified methodology that combines these paradigms for the specifics of EDMS constitutes a research gap. The aim of this study is to develop and experimentally substantiate

***Corresponding author: Assel Omarbekova (omarbekova_as@enu.kz), Aizhan Nazyrova (nazyrova_aye_1@enu.kz), Zhanar Lamasheva (lamasheva_zhb@enu.kz)**
DOI: https://doi.org/10.47738/jads.v7i3.1492

an intelligent system for the quantitative, reproducible, and explainable assessment of EDMS cybersecurity. The scientific novelty consists of a domain ontology of EDMS assets and threats with confidentiality, integrity, and availability (CIA) weights; a hybrid model in which fuzzy inference, a machine-learning (ML) ensemble, and a Bayesian network complement one another; a composite index (CSAI) with adaptive re-weighting based on the fuzzy analytic hierarchy process (Fuzzy-AHP); and a built-in explainability layer. The remainder of the paper is organized as follows. Section 2 reviews related work; Section 3 presents the materials and methods; Section 4 reports the results; Section 5 provides a comparative analysis; Section 6 discusses implications and limitations; and Section 7 concludes.

2. Related work

Classical risk-assessment methodologies formalize risk as a function of probability and impact, relying on standardized severity scales and expected-loss metrics [3], [4], [13], [14]. Although methodologically mature [5], [15], they are predominantly periodic and qualitative and poorly capture asset interdependence and the cascading nature of threat realization in EDMS.

Fuzzy logic formalizes the linguistic uncertainty of expert judgements [6]: Mamdani and Takagi–Sugeno inference systems [7] yield interpretable estimates, yet their rule base is static and does not learn from data. Machine-learning methods - tree ensembles [8], [9] and neural networks [16] - achieve high accuracy in recognizing risky states [10], [17] but are sensitive to data scarcity and class imbalance and are limited in interpretability. Bayesian networks and attack graphs model causal dependencies and multi-step scenarios [11], [12], [18], [19] but require prior distributions and calibration.

Multi-criteria decision-making methods, in particular the fuzzy analytic hierarchy process with consistency checking [20], [21], are used to aggregate heterogeneous estimates. Existing hybrid solutions typically realize a sequential composition of two methods and rarely provide consistent domain-level aggregation, weight adaptation, and end-to-end explainability [22], [23]. EDMS, meanwhile, are document-centric, with fine-grained context-dependent access and intensive integrations, while regulatory requirements shift the focus toward continuous compliance control and a measurable estimate of residual risk. Comprehensive solutions integrating fuzzy inference, machine learning, and Bayesian modelling with a domain ontology and an explainability layer for EDMS are fragmentary; the present work fills this gap.

A more detailed examination of recent hybrid cybersecurity-assessment frameworks helps to position the present contribution. A first family of methods couples fuzzy inference with machine learning: an adaptive neuro-fuzzy inference system (ANFIS) embeds a fuzzy rule base inside a trainable network so that membership-function parameters are learned from data rather than fixed by experts [24], while related studies use fuzzy logic to post-process classifier outputs and recover interpretability lost in purely data-driven models. A second family integrates probabilistic graphical models with attack modelling, combining Bayesian networks with automatically generated attack graphs to estimate the likelihood of multi-step intrusions and to propagate evidence across causally linked states [12], [18], [19], [25]. A third family applies multi-criteria decision analysis, most commonly the fuzzy analytic hierarchy process, to fuse heterogeneous expert and data-driven indicators into a single comparable score [20], [21]. Despite their individual strengths, these frameworks share three recurring limitations: they combine only two paradigms in a fixed, sequential pipeline; they rely on static aggregation weights that are not re-estimated as the environment or the incident history evolves; and they seldom couple a domain ontology with an end-to-end explainability layer. Moreover, none is tailored to the document-centric, integration-intensive setting of EDMS, in which fine-grained, context-dependent access control and continuous-compliance obligations demand both data-driven adaptation and causal, auditable reasoning. The framework proposed here departs from this body of work by jointly integrating three complementary paradigms - Mamdani fuzzy inference, a machine-learning stacking ensemble, and a Bayesian threat network with an attack graph - under a shared EDMS asset-and-threat ontology, with adaptive Fuzzy-AHP re-weighting driven by confirmed incidents and a unified, two-level explainability layer.

3. Materials and Methods

3.1. Problem statement and threat model

An EDMS is treated as a collection of assets, threats, and protective controls. Each asset is characterized by its criticality with respect to confidentiality, integrity, and availability, aggregated with domain weights:

$$C_a = w_C C_a^{conf} + w_I C_a^{int} + w_A C_a^{avail},\ w_C + w_I + w_A = 1 \quad (1)$$

The base risk of an asset–threat pair is the product of the threat realization probability, the vulnerability, and the impact:

$$R_{a,t} = P(t) \cdot V_{a,t} \cdot I_{a,t} \quad (2)$$

For a monetary interpretation and prioritization, the annual loss expectancy metric is used:

$$ALE = SLE \times ARO = (AV \cdot EF) \times ARO \quad (3)$$

Partial risks in an EDMS are not independent: shared components, shared accounts, and end-to-end integrations induce correlations and cascading effects; therefore, aggregation is performed by a non-linear weighted convolution, and multi-step scenarios are modelled by a Bayesian network. The domain ontology of EDMS assets and threats (table 1) provides contextual weights and links features to domain entities.

Table 1. Domain ontology of electronic document management system assets and threats

Asset class	Dominant CIA	Characteristic threats	Key controls
Document repository	Confidentiality	Unauthorized access, leakage, exfiltration	Encryption, access control, DLP
Versions and metadata	Integrity	Version tampering, falsification, log deletion	Hashing, immutable logs, e-signature
Routing service	Availability	Denial of service, queue overload	Redundancy, rate limiting, monitoring
Access-control subsystem	Confidentiality / Integrity	Privilege escalation, access-control bypass	RBAC/ABAC, least privilege
Integration gateways (API)	Confidentiality	Leakage via weak integrations, injections	Token authentication, validation, isolation
Audit logs	Integrity / Availability	Trace hiding, event deletion	Centralization, tamper protection, retention
Cloud / SaaS services	Confidentiality / Availability	Misconfiguration, insecure object storage, account hijacking, data residency violation	CSPM, encryption at rest/in transit, hardened IAM
Mobile access endpoints	Confidentiality	Device loss or theft, insecure sessions, mobile malware, screen capture	MDM/MAM, device attestation, session binding, DLP
Identity and zero-trust layer	Confidentiality / Integrity	Token theft, lateral movement, trust misconfiguration, MFA bypass	Continuous verification, micro-segmentation, phishing-resistant MFA
Third-party / supply-chain integrations	Confidentiality / Integrity	Compromised connector, over-privileged scopes, malicious update	Vendor risk assessment, scoped tokens, SBOM, egress monitoring

3.2. Architecture of the intelligent system

The system has a six-layer architecture: data collection and normalization, mapping to the ontology, an analytical core of three subsystems (fuzzy, ensemble, Bayesian), an aggregation and index layer, and a presentation layer with explanations. A feedback loop provides re-training and re-weighting (figure 1).

 ISSN 2723-6471

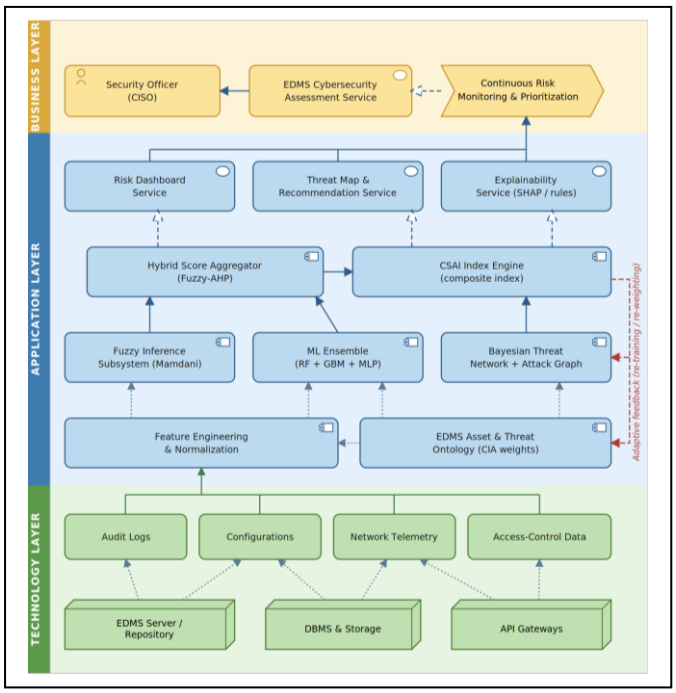

Figure 1. Multi-layer architecture of the intelligent EDMS cybersecurity assessment system

The methodological pipeline (figure 2) comprises collection and normalization, assignment of CIA weights, fuzzification, computation of probabilities by the ensemble and the Bayesian network, Fuzzy-AHP aggregation, and the formation of the CSAI index with explanations.

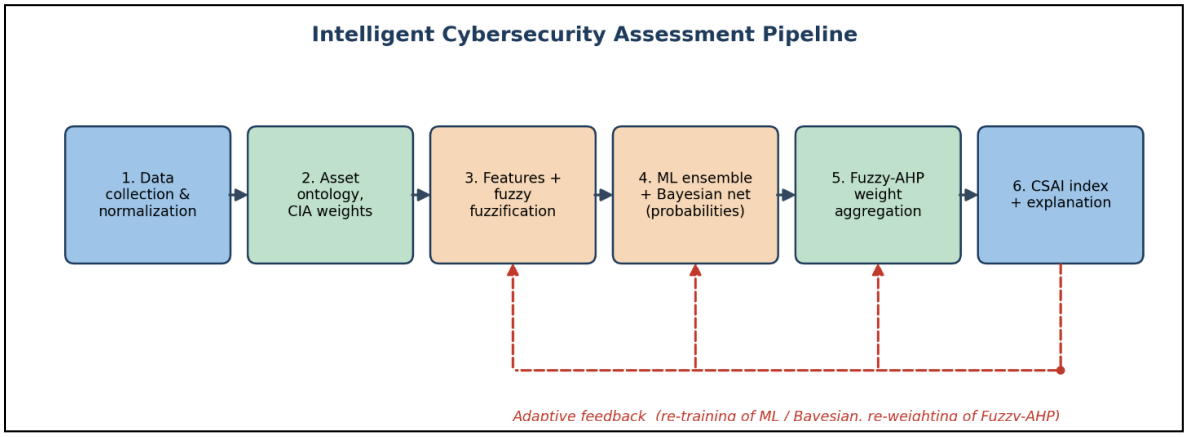

Figure 2. Intelligent cybersecurity assessment pipeline with adaptive feedback

3.3. Fuzzy assessment subsystem

The input linguistic variables (vulnerability severity, asset criticality, control maturity, anomalous activity) are described by fuzzy terms [6]. The triangular membership function is:

$$\mu_{tri}(x;\ a,b,c)\ =\ max(min(\frac{x-a}{b-a},\frac{c-x}{c-b}),0) \tag{4}$$

Trapezoidal membership functions are used for the boundary terms:

$$\mu_{trap}(x;\ a,b,c,d)\ =\ max(min(\frac{x-a}{b-a},1,\frac{d-x}{d-c}),0) \tag{5}$$

The resulting shapes of these membership functions for the vulnerability-severity and risk-level variables are illustrated in figure 3.

ISSN 2723-6471

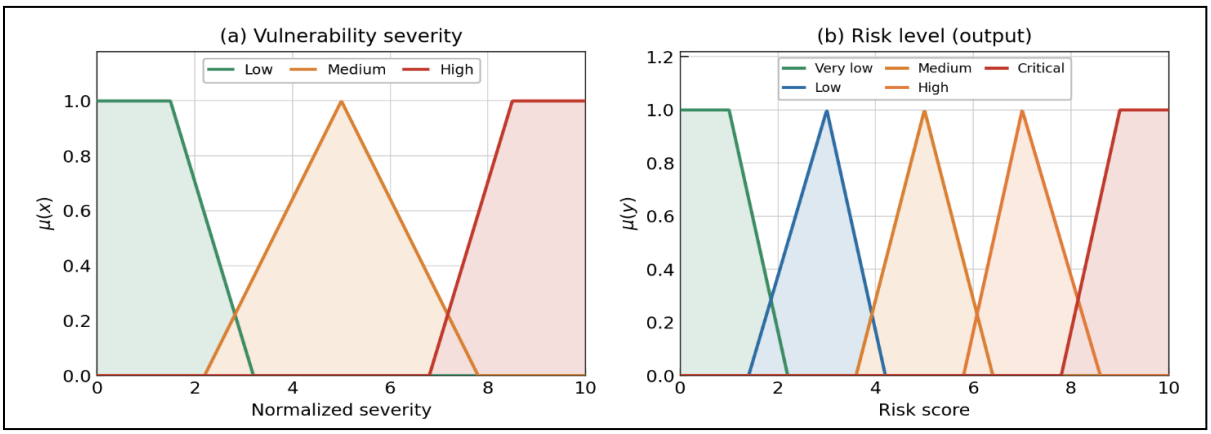

Figure 3. Membership functions: (a) vulnerability severity; (b) risk level

The firing strength of a rule (minimum t-norm) is:

$$\alpha_k = min_j \, \mu_{A_{kj}}(x_j) \tag{6}$$

Aggregation by the Mamdani method [7]:

$$\mu_C(y) = max_k \, min(\alpha_k, \mu_{C_k}(y)) \tag{7}$$

Defuzzification by the centroid method:

$$y^* = \frac{\int y \cdot \mu_c(y) \, dy}{\int \mu_c(y) \, dy} \tag{8}$$

The knowledge base contains about eighty production rules; the membership-function parameters are calibrated on the training sample by coordinate descent. The linguistic variables are listed in table 3.

To make the knowledge base transparent and reproducible, representative production rules are listed in table 2. Each rule maps the input linguistic variables of table 3 (vulnerability severity, asset criticality, control maturity, and anomalous activity) to the output risk level; the full base of about eighty rules follows the same structure and is available in the supplementary material.

Table 2. Representative production rules of the fuzzy knowledge base

Rule	Vulnerability severity	Asset criticality	Control maturity	Anomalous activity	Risk level
R1	High	—	Weak	—	**Critical**
R2	High	High	Basic	—	**High**
R3	Medium	—	—	High	**High**
R4	—	High	not Mature	High	**High**
R5	Medium	—	Mature	Low	**Medium**
R6	Low	—	Mature	Low	**Very low**
R7	Low	Low	—	—	**Low**
R8	—	—	Weak	High	**Critical**

Table 3. Linguistic variables of the fuzzy subsystem

Variable	Terms	Support / scale
Vulnerability severity (input)	Low, Medium, High	[0; 10], normalized severity scale
Asset criticality (input)	Low, Medium, High	[0; 1], CIA convolution
Control maturity (input)	Weak, Basic, Mature	[0; 1], control coverage index
Anomalous activity (input)	Low, Moderate, High	[0; 1], normalized frequency
Risk level (output)	Very low, Low, Medium, High, Critical	[0; 10], risk score

 ISSN 2723-6471

3.4. Bayesian threat network and attack graph

The causal structure of threat development is modelled by a Bayesian network [11], [12]: root causes, intermediate states, and target events (figure 4). The full network contains 23 nodes and 41 directed edges; each node takes discrete states (binary "realized / not realized" for most, three levels for severity nodes). The attack graph is generated automatically from the environment model and the vulnerability catalogue [18], [19] by traversing reachable states with respect to exploitation pre- and post-conditions.

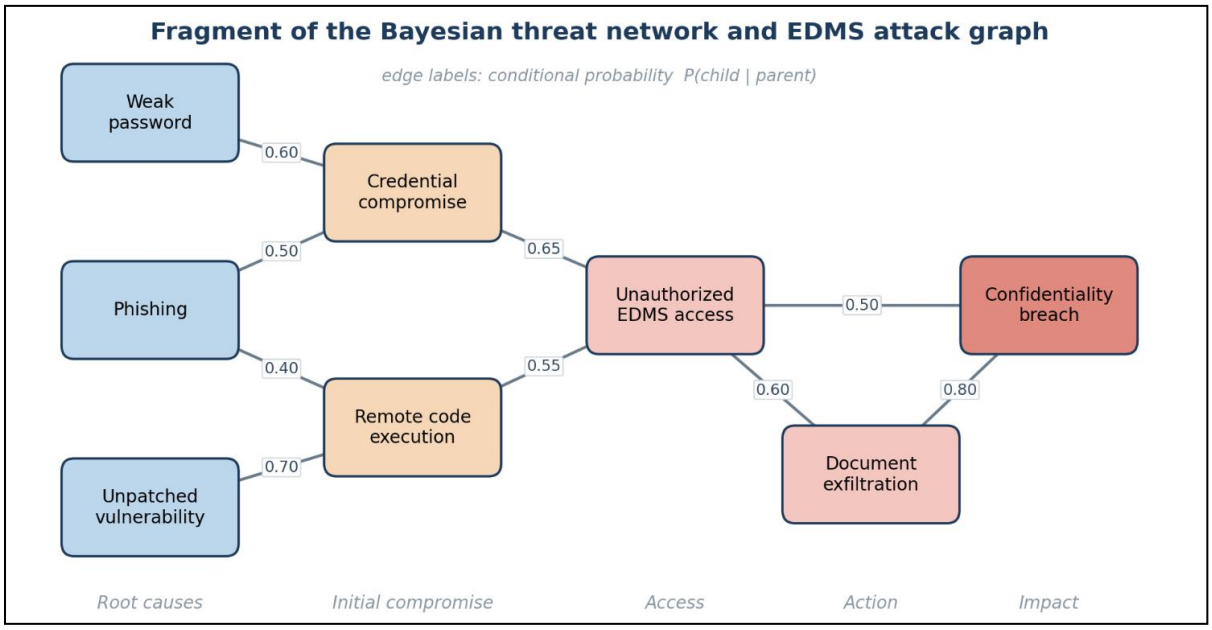

Figure 4. Fragment of the Bayesian threat network and EDMS attack graph

For full reproducibility, the complete node and edge specification of the Bayesian threat network is given below (the corresponding graph is also provided as supplementary material). The network is a directed acyclic graph organized into three layers - root causes, intermediate states, and target events.

Root-cause nodes (8): N1 Unpatched vulnerability; N2 Weak credential policy; N3 Misconfigured access control; N4 Insecure integration/API; N5 Insufficient monitoring; N6 Phishing exposure; N7 Weak network segmentation; N8 Insider risk.

Intermediate-state nodes (10): N9 Credential compromise; N10 Initial foothold; N11 Privilege escalation; N12 Lateral movement; N13 Repository access; N14 Version/integrity tampering; N15 Integration-gateway breach; N16 Audit-log tampering; N17 Data staged for exfiltration; N18 Detection evasion.

Target-event nodes (5): N19 Confidentiality breach / data leakage; N20 Integrity violation; N21 Availability disruption; N22 Compliance/regulatory violation; N23 Full repository compromise. Most nodes are binary (realized / not realized); severity-related nodes (N11, N14, N19) take three levels.

Directed edges (41): N6→N9, N2→N9, N1→N10, N4→N15, N3→N11, N7→N12, N8→N13, N5→N18, N9→N10, N9→N11, N10→N11, N10→N12, N11→N12, N11→N13, N12→N13, N12→N15, N13→N17, N13→N14, N15→N17, N15→N13, N4→N17, N14→N20, N13→N16, N11→N16, N16→N18, N5→N16, N18→N19, N18→N21, N17→N19, N13→N19, N14→N23, N13→N23, N15→N23, N7→N21, N1→N21, N23→N22, N19→N22, N20→N22, N21→N22, N23→N20, N23→N19.

Conditional probability tables for multi-parent nodes (N11, N12, N13, N18) use a Noisy-OR parameterization, which reduces the number of elicited parameters from exponential to linear in the number of parents; the remaining tables were calibrated on empirical data as described below. Upon arrival of evidence, the posterior probability of a hypothesis is computed by Bayes' rule:

$$P(\theta_i \mid E) = \frac{P(E|\theta_i)\, P(\theta_i)}{\sum_j P(E|\theta_j)\, P(\theta_j)} \quad (9)$$

The joint distribution is factorized according to the network structure:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i)) \tag{10}$$

The probability of a multi-step scenario and the system risk are:

$$P_{path} = \prod_{e \in path} p_e, \quad R_{sys} = 1 - \prod_p (1 - P_p) \tag{11}$$

The conditional probability tables (CPTs) were initialized by experts: five EDMS information-security specialists (6–15 years of experience) independently rated the conditional probabilities on a structured scale, after which the estimates were aggregated by averaging with a consistency check. For nodes with many parents, a Noisy-OR parametric model was used to reduce the number of elicited parameters. The CPTs were then calibrated on empirical data by maximum likelihood with Laplace smoothing; approximate sampling inference was used for large fragments.

Several measures were taken to limit the influence of the small expert panel on the conditional probability tables and to improve their stability. First, the five specialists rated probabilities independently and blind to one another, after which their estimates were aggregated and screened for consistency, reducing individual bias. Second, the Noisy-OR parameterization sharply lowers the number of quantities that must be elicited for multi-parent nodes, so that fewer expert judgements drive the model. Third, and most importantly, the elicited values were used only as informative priors and were subsequently re-estimated on the empirical data by maximum-likelihood with Laplace smoothing; this data-driven calibration progressively dominates the priors as evidence accumulates, so the final tables depend far less on the initial expert ratings than a purely subjective network would. A leave-one-expert-out sensitivity check produced only minor changes in the posterior risk estimates (mean absolute change in CSAI below two points), indicating that the network is robust to the composition of the panel. Broadening the expert pool would nonetheless further strengthen the elicitation, and this is identified as a direction for future work.

3.5. Machine-learning ensemble and feature engineering

A baseline logistic model defines the compromise probability:

$$P(y = 1 \mid x) = \frac{1}{1 + exp(-(w^T x + b))} \tag{12}$$

The final estimate is formed by stacking [26] a random forest [8], gradient boosting [9], and a multilayer perceptron [16]:

$$\hat{\rho}(c \mid x) = \sum_{m=1}^{M} w_m f_m(x), \quad \sum_m w_m = 1, \; w_m \geq 0 \tag{13}$$

Feature selection uses information gain through Shannon entropy, with importance additionally assessed by SHAP [22]:

$$IG(S, f) = H(S) - \sum_{v \in V_f} \frac{|S_v|}{|S|} H(S_v), \quad H(S) = -\sum_c p_c log_2 p_c \tag{14}$$

Training minimizes the regularized empirical risk to prevent overfitting:

$$min_\Theta \; \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(y_i, \hat{y}_i) + \lambda ||\Theta||^2_2 \tag{15}$$

Class imbalance is mitigated by class weighting and synthetic minority over-sampling (SMOTE) [27] performed strictly within the training folds. The feature groups and data sources are given in table 4.

Table 4. Feature groups and data sources

Group	Example features	Type	Source
Configuration	Age of unpatched vulnerabilities, password-policy strength, backup frequency	Numeric / ordinal	Configurations, scanners
Access control	Access-rights entropy, share of privileged accounts, expired sessions	Numeric	Access-control subsystem
Behavioural	Anomalous-login frequency, deviation from access profile, download spikes	Numeric	Audit logs

Architectural	Network segmentation, share of external integrations, document encryption coverage	Numeric / binary	Network telemetry, inventory
Monitoring	Logging depth, detection coverage, response time	Numeric / ordinal	Monitoring tools

3.6. Hybrid aggregation and the CSAI index

The partial estimates of the three subsystems are aggregated across security domains; the domain weights are determined by the fuzzy analytic hierarchy process [20], [21] with consistency control:

$$CI = \frac{\lambda_{max} - n}{n-1}, \quad CR = \frac{CI}{RI} < 0.1 \tag{16}$$

Weights are extracted from triangular fuzzy numbers by centroid defuzzification:

$$w_i = \frac{\frac{l_i + m_i + u_i}{3}}{\sum_k \frac{l_k + m_k + u_k}{3}} \tag{17}$$

The domain risk is normalized with respect to asset criticality:

$$\hat{R}_d = \frac{\sum_{a,t} C_a R_{a,t}}{\sum_{a,t} C_a R^{max}{}_{a,t}} \tag{18}$$

The composite cybersecurity assessment index CSAI (a 0–100 scale, with higher values denoting stronger protection) is:

$$CSAI = 100 \cdot (1 - \sum_{d=1}^{D} W_d \hat{R}_d), \quad \hat{R}_d \in [0,1] \tag{19}$$

The choice of a weighted additive convolution of normalized domain risks is grounded in multi-attribute utility theory: under mutual preferential independence of criteria, the value function admits an additive form [28], [29]. Domain risks are first normalized to the interval [0, 1] (Eq. 18), which removes scale differences and makes the weights comparable, while cascading asset dependencies are captured not at the convolution level but within the domain estimates through the Bayesian network. The form that is linear in the normalized risks ensures monotonicity, transparency, and additive decomposability of the domain contributions, which is important for explainability; complementation to unity and scaling by 100 translate the convolution into a familiar percentage maturity scale aligned with the five-level linguistic classification. The weights are obtained by the Fuzzy-AHP procedure with a consistency check, which distinguishes the index from heuristic additive scales.

Two alternative aggregation schemes were also considered and explicitly rejected. A multiplicative utility function of the form $CSAI \propto \prod d\ (1 - \hat{R}_d)\hat{}W_d$, and the closely related weighted geometric mean, impose a conjunctive, non-compensatory logic in which a single severely compromised domain drives the whole index toward its worst value irrespective of the others. Although this behaviour is attractive when criteria must not compensate for one another, it has three drawbacks in the EDMS setting. First, the multiplicative form collapses to zero whenever any normalized domain risk approaches its upper bound ($\hat{R}d \to 1$), producing saturation-prone scores that are unstable and difficult to read on a continuous maturity scale. Second, the logarithmic transformation needed to estimate its weights destroys the additive decomposability on which the explainability layer relies: the contribution of each domain to the final score then depends non-linearly on the values of all other domains, obscuring the per-domain attribution that security officers use to prioritize remediation. Third, multiplicative aggregation amplifies the influence of noisy or weakly observed domains, reducing robustness on operational data. By contrast, the additive convolution satisfies mutual preferential independence of the (already Bayesian-coupled) domain risks, yields transparent and monotone per-domain contributions, and maps naturally onto the five-level linguistic scale. Crucially, conjunctive and cascading effects between assets are not lost under the additive form: they are modelled upstream, within each domain estimate, by the Bayesian threat network, so the domain-level convolution can remain additive without sacrificing the representation of non-independent threats. The weighted additive convolution is therefore adopted as the principled choice for the CSAI index, while multiplicative and geometric-mean variants remain useful baselines for sensitivity analysis.

Interpretation is given by a five-level scale (table 5); adaptive feedback adjusts the domain weights according to confirmed incidents.

 ISSN 2723-6471

Table 5. Interpretation scale of the composite cybersecurity assessment index

CSAI range	Protection level	Recommended response
85–100	Very high	Maintain regime, routine monitoring
70–84	High	Targeted strengthening of weak domains
50–69	Medium	Prioritized corrective-action plan
30–49	Low	Immediate remediation of critical gaps
0–29	Critical	Emergency regime, escalation, asset isolation

3.7. System operation algorithm

Algorithm 1 summarizes the runtime procedure described in Sections 3.1–3.2: it takes a configuration state as input, computes the fuzzy, ML, and Bayesian risk estimates in parallel (steps 2–5), aggregates them into the CSAI index via Fuzzy-AHP weighting (steps 6–7), and generates an explanation while updating the domain weights from feedback (step 8).

Algorithm 1. Hybrid EDMS cybersecurity assessment

Input: state s, ontology O, knowledge bases (rules, network, models), domain weights W.

1. Extract and normalize features x ← Features(s); map assets to O.
2. Fuzzification: compute membership degrees μ(x).
3. Fuzzy inference: Eq. (6)–(8) → r_fuzzy.
4. ML ensemble: compromise probability p_ml via Eq. (12)–(13).
5. Bayesian net: Eq. (9)–(11) → r_bayes.
6. Normalize domain risks Eq. (18), aggregate Fuzzy-AHP weights Eq. (16)–(17).
7. Compute CSAI via Eq. (19), determine level from table 5.
8. Generate explanation and update W via feedback.

Output: CSAI, domain-risk profile, ranked recommendations, explanation.

3.8.Dataset, metrics, and experimental setup

Owing to the confidentiality requirements of operational EDMS environments, the dataset was derived from the operational electronic document management system of L.N. Gumilyov Eurasian National University (Astana, Kazakhstan). After anonymization and preprocessing to remove all personally identifiable and organization-specific information, the final dataset comprised 10,239 configuration states described by 42 cybersecurity-related features. The collected records included configuration parameters, access-control information, audit-log indicators, network telemetry, and monitoring metrics. The 42 features were derived from five groups of cybersecurity indicators: configuration management, access control, behavioural monitoring, architectural characteristics, and monitoring effectiveness (table 4). All personally identifiable and organization-specific information was removed before analysis in accordance with data-protection requirements.

The target risk level was assigned by three independent cybersecurity experts using a five-level risk scale. Final labels were determined by majority voting, and inter-rater agreement was substantial (Cohen's $\kappa = 0.81$). The dataset was divided into training (approximately 70%; 7,166 states), validation (approximately 15%; 1,536 states), and test (approximately 15%; 1,537 states) subsets using stratified sampling.

To make the risk categories reproducible, the experts applied a written labeling rubric rather than an unaided subjective judgement. Each configuration state was rated along four criteria - the exploitability of the vulnerabilities present, the exposure and attack surface, the coverage and maturity of compensating controls, and the potential CIA impact on the affected assets. Each criterion was scored on a four-point scale (0–3): exploitability (0 = no exploitable vulnerabilities, 1 = low-severity vulnerabilities requiring complex exploitation, 2 = exploitable vulnerabilities with moderate prerequisites, 3 = easily exploitable or publicly weaponized vulnerabilities); exposure (0 = isolated internal assets with

minimal attack surface, 1 = limited internal exposure, 2 = externally reachable services or moderate attack surface, 3 = Internet-facing critical services or extensive exposure); control maturity (0 = mature preventive and detective controls fully implemented, 1 = mostly effective controls with minor gaps, 2 = partial control coverage, 3 = weak or absent controls); and CIA impact (0 = negligible operational impact, 1 = limited impact on non-critical assets, 2 = significant impact on important assets, 3 = severe impact on critical assets affecting confidentiality, integrity, or availability). The criterion scores were summed to obtain a total risk score ranging from 0 to 12. Risk labels were assigned using fixed thresholds: Very low (0–2), Low (3–5), Medium (6–8), High (9–10), and Critical (11–12). Disagreements were resolved by majority vote, and the substantial inter-rater agreement (Cohen's $\kappa = 0.81$) indicates that the criteria were applied consistently. The resulting distribution of risk classes across the training, validation, and test subsets is summarized in table 6.

Table 6. Dataset characteristics and sample split

Risk class	Share, %	Training	Validation	Test
Very low	19.5	1,400	300	300
Low	21.1	1,510	324	324
Medium	24.0	1,720	369	369
High	20.6	1,476	316	316
Critical	14.8	1,060	227	228

Quality was measured by accuracy, recall, the F1 score, the Matthews correlation coefficient [30], and the area under the ROC curve [31]:

$$F_1 = \frac{2PR}{P+R}, \quad MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{20}$$

Hyperparameters were tuned by grid search with five-fold cross-validation; the final values and the software environment are given in table 7. Experiments were run on Python 3.11 using scikit-learn 1.4 [32], XGBoost 2.0, LightGBM 4.3, CatBoost 1.2, TensorFlow/Keras 2.15 (multilayer perceptron), pgmpy 0.1.25 (Bayesian network), and scikit-fuzzy 0.4 (fuzzy subsystem). The significance of differences between models was tested using the non-parametric Wilcoxon signed-rank test across folds. Results are reported as mean ± standard deviation across the five cross-validation folds, and approximate 95% confidence intervals were computed as mean ± 1.96 × SD. The random-number generator seed was fixed (seed = 42) to ensure reproducibility.

For computational reproducibility, the hardware environment is reported alongside the software stack of table 7. All experiments were run on an Intel Core i7-12700H (14 physical cores, 2.30 GHz), 32 GB RAM, and NVIDIA GeForce RTX 4060 (8 GB) under Windows 11 Pro 23H2 (64-bit). End-to-end training of the full hybrid model required approximately 36 minutes, and inference for a single EDMS state-fuzzy inference, ensemble prediction, Bayesian propagation, and CSAI aggregation-completed in about 17 ms, which supports near-real-time assessment. Regarding statistical significance, the Wilcoxon signed-rank test across folds confirms that the hybrid model improves on every baseline; exact two-sided p-values are: versus the stacking ensemble $p = 0.031$; versus CatBoost $p = 0.018$; versus LightGBM $p = 0.014$; versus XGBoost $p = 0.009$; and $p < 0.001$ against all remaining models.

Because the reported metrics are high, the risk of overfitting to the institutional dataset was explicitly examined. All model selection, feature engineering, and SMOTE over-sampling were performed strictly inside the training folds of a nested cross-validation, so the held-out test set (1,537 states) never influenced training or hyperparameter tuning. The gap between training and test F1 for the hybrid model was small (training F1 = 0.961 versus test F1 = 0.946, a difference of 1.5 points), and the per-fold standard deviation was low (±0.006), which is inconsistent with severe overfitting. L2 regularization, dropout in the multilayer perceptron, and out-of-fold meta-model training further constrain model capacity, and the Wilcoxon signed-rank test confirms that the improvement over the strongest baseline is statistically significant rather than an artefact of a favourable split. These analyses nevertheless establish only internal validity: they cannot substitute for evaluation on data from other organizations. External validation on independent industrial, governmental, and financial EDMS deployments - whose threat profiles may differ substantially from a university environment - is therefore a priority for future work, and the generalizability claims have been tempered accordingly (see Limitations).

Table 7. Final model hyperparameters and software environment

Model / component	Final hyperparameters	Environment
Random forest	n_estimators=500, max_depth=16, min_samples_leaf=4, max_features=sqrt	scikit-learn 1.4
Gradient boosting (XGBoost)	lr=0.05, n_estimators=600, max_depth=6, subsample=0.8, colsample=0.8	XGBoost 2.0
LightGBM	lr=0.05, num_leaves=63, n_estimators=600, feature_fraction=0.8	LightGBM 4.3
CatBoost	lr=0.05, depth=6, iterations=600, l2_leaf_reg=3	CatBoost 1.2
Multilayer perceptron	layers 128–64, ReLU, dropout=0.3, L2=1e-3, Adam, batch=64	Keras 2.15
Meta-model (stacking)	logistic regression, L2=0.1, out-of-fold training	scikit-learn 1.4
ANFIS (baseline)	3 terms per input, hybrid learning, 200 epochs	custom implementation
Bayesian network	23 nodes, 41 edges, Noisy-OR, Laplace smoothing	pgmpy 0.1.25
Fuzzy-AHP / fuzzy	Saaty scale 1–9, CR < 0.1; ~80 rules, centroid	scikit-fuzzy 0.4

4. Results

4.1. Comparative model performance

The proposed hybrid model was compared with single classical models and with modern gradient-boosting methods (XGBoost [9], LightGBM [33], CatBoost [34]) and the ANFIS [24] under an identical split and unified metrics. Summary results, reported as the mean across the five cross-validation folds with the standard deviation and the approximate 95% confidence interval for the F1 score, are presented in table 8. The hybrid model attains the best values on all metrics: F1 = 0.946, MCC = 0.927, AUC = 0.972. The F1 gain is 2.4 percentage points relative to the strongest baseline (the stacking ensemble) and up to 10.4 percentage points relative to logistic regression.

Table 8. Comparison of risk-assessment model performance on the test set

Model	Acc	F1 (M ± SD)	MCC	AUC	95% CI (F1)
Logistic regression	0.851	0.842 ± 0.011	0.805	0.864	[0.821; 0.863]
Fuzzy (Mamdani)	0.889	0.881 ± 0.012	0.851	0.901	[0.858; 0.904]
ANFIS	0.901	0.899 ± 0.010	0.869	0.918	[0.880; 0.918]
Random forest	0.907	0.903 ± 0.009	0.879	0.928	[0.886; 0.920]
XGBoost	0.918	0.915 ± 0.008	0.893	0.940	[0.900; 0.930]
LightGBM	0.920	0.917 ± 0.008	0.896	0.943	[0.902; 0.932]
CatBoost	0.922	0.919 ± 0.008	0.898	0.945	[0.904; 0.934]
Stacking ensemble	0.925	0.922 ± 0.007	0.902	0.951	[0.909; 0.935]
Hybrid (proposed)	0.948	0.946 ± 0.006	0.927	0.972	[0.935; 0.957]

The discriminative ability of the models is confirmed by ROC analysis [31] (figure 5). The curve of the proposed model dominates the others across the entire range of thresholds, indicating a stable advantage under different sensitivity–specificity trade-offs.

ISSN 2723-6471

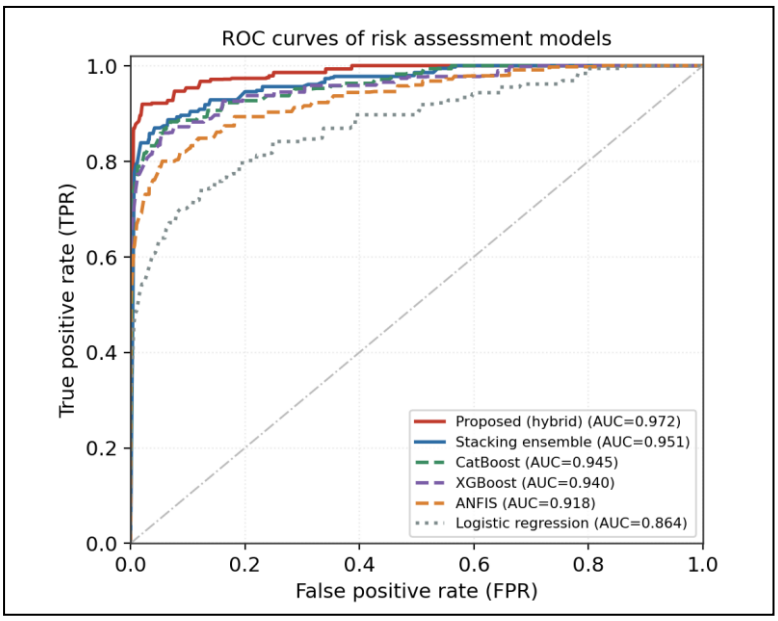

Figure 5. ROC curves of the compared risk-assessment models

4.2. Error analysis

The confusion matrix of the hybrid model (figure 6) shows high diagonal concentration: the share of correct decisions per class ranges from 93% to 96%. The predominant errors are adjacent shifts between neighbouring risk levels (for example, medium–high), which is consistent with the ordinal nature of the target scale and does not lead to gross misclassifications between extreme classes. This property is important in practice, since adjacent errors have little effect on the prioritization of corrective measures.

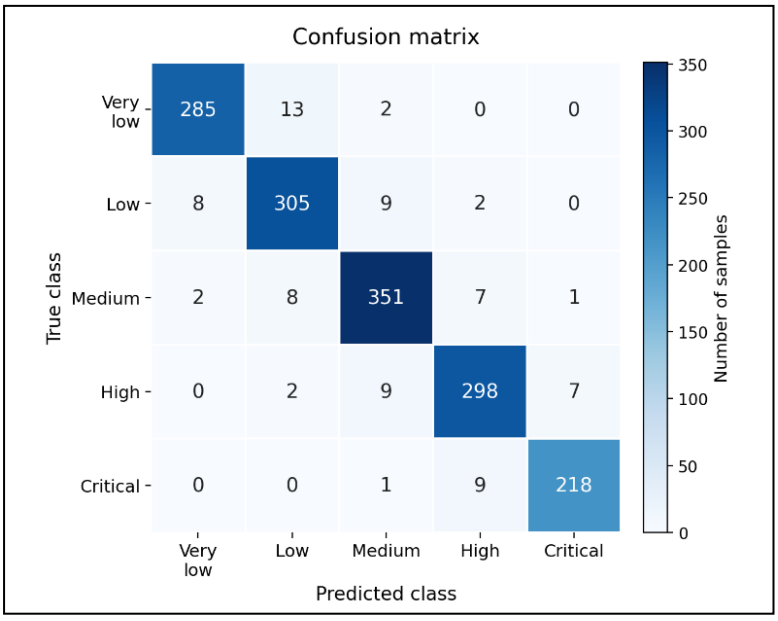

Figure 6. Confusion matrix (absolute counts) of the hybrid model on the test set

4.3. Feature importance and explainability

Feature-importance analysis, performed using mean absolute SHAP values [22] and impurity reduction, revealed the dominant contribution of vulnerability management and access control (figure 7). The most significant features are the age of unpatched vulnerabilities and access-rights entropy, followed by anomalous-login frequency and document encryption coverage. The agreement of the feature ranking with expert views on EDMS protection priorities confirms the substantive validity of the model and increases confidence in its recommendations.

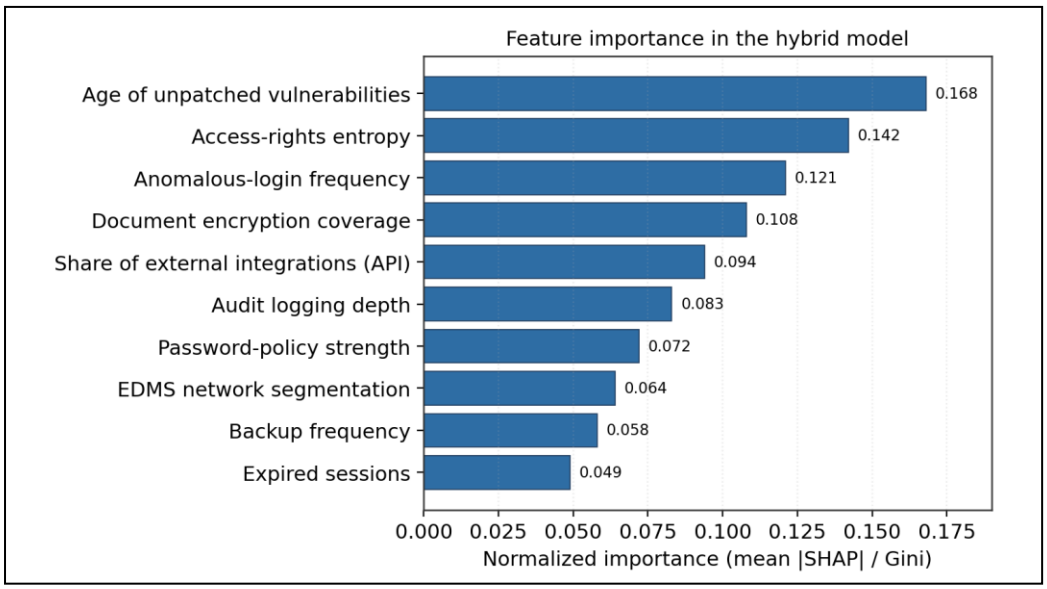

Figure 7. Feature importance in the hybrid model (normalized values)

Explainability is realized at two levels. The global level characterizes the model's overall behaviour through averaged feature importance and the domain decomposition of the index, while the local level explains an individual estimate: for a specific EDMS state, a list of the most influential features is formed with the direction and magnitude of their contribution, supplemented by the fired fuzzy rules and the most probable attack paths from the Bayesian subsystem.

To make the explainability layer concrete, a single end-to-end example is traced for one representative test state. The state received CSAI = 54 and was classified as Medium protection level (table 6). At the local level, the largest feature contributions, ranked by signed SHAP value, were: age of unpatched vulnerabilities (SHAP = +0.168, pushing toward higher risk), access-rights entropy (+0.142), anomalous-login frequency (+0.121), and low document-encryption coverage (+0.108); the main risk-reducing factor was backup frequency (−0.058). In the fuzzy subsystem, the rules with the highest firing strength were R1 and R4, reflecting high vulnerability severity combined with weak control maturity. In the Bayesian subsystem, the most probable attack path was N1 → N10 → N11 → N13 → N19 (unpatched vulnerability → initial foothold → privilege escalation → repository access → confidentiality breach), with a posterior probability of 0.82. Combining these signals, the system produced the ranked recommendations: (1) patch the identified vulnerabilities, (2) reduce access-rights entropy by applying the principle of least privilege, (3) strengthen monitoring of anomalous logins, and (4) increase document-encryption coverage. The expected effect of implementing these measures was an increase of CSAI from 54 to 84. This trace shows how a single score is decomposed into feature-, rule-, and path-level evidence that an analyst can act on.

4.4. Ablation analysis

An ablation experiment was conducted to assess the contribution of each component (table 9, figure 8). The baseline configuration is the ML stacking ensemble (F1 = 0.922, as in table 8). Sequentially adding the fuzzy subsystem and the Bayesian network monotonically improves quality: the F1 score increases from 0.922 to 0.946 and the Matthews coefficient from 0.902 to 0.927. This confirms that the components do not duplicate but complement one another: fuzzy inference brings expert interpretability, and the Bayesian network adds causal sensitivity to multi-step scenarios.

Table 9. Ablation analysis of the hybrid model components

Configuration	F1	MCC	AUC	Δ F1
ML only (stacking)	0.922	0.902	0.951	-
ML + fuzzy subsystem	0.931	0.911	0.958	+0.009
ML + Bayesian network	0.936	0.916	0.962	+0.014
Full hybrid model	0.946	0.927	0.972	+0.024

 ISSN 2723-6471

The learning curve (figure 8) shows a narrowing gap between training and validation quality as the data volume grows and a plateau after about eight thousand examples, which indicates the absence of substantial overfitting and the sufficiency of the sample for stable generalization.

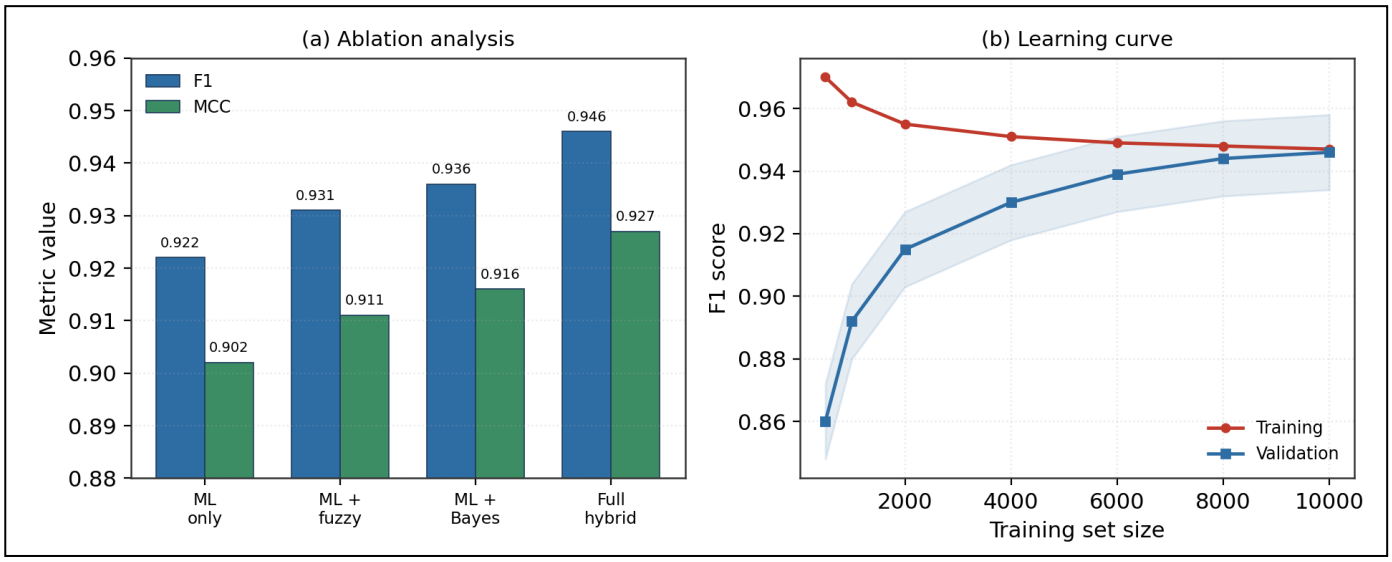

Figure 8. (a) Ablation analysis of components; (b) learning curve of the hybrid model

4.5. Security profile and the CSAI index

Applying the system to a reference EDMS deployment produced a security-maturity profile by domain (figure 9). The initial state was characterized by gaps in vulnerability management, auditing, and incident response, reflected in a CSAI value of 54 (medium protection level). After implementing the measures prioritized by the system and ranked by expected risk reduction, the profile levelled out and the CSAI rose to 84 (high level).

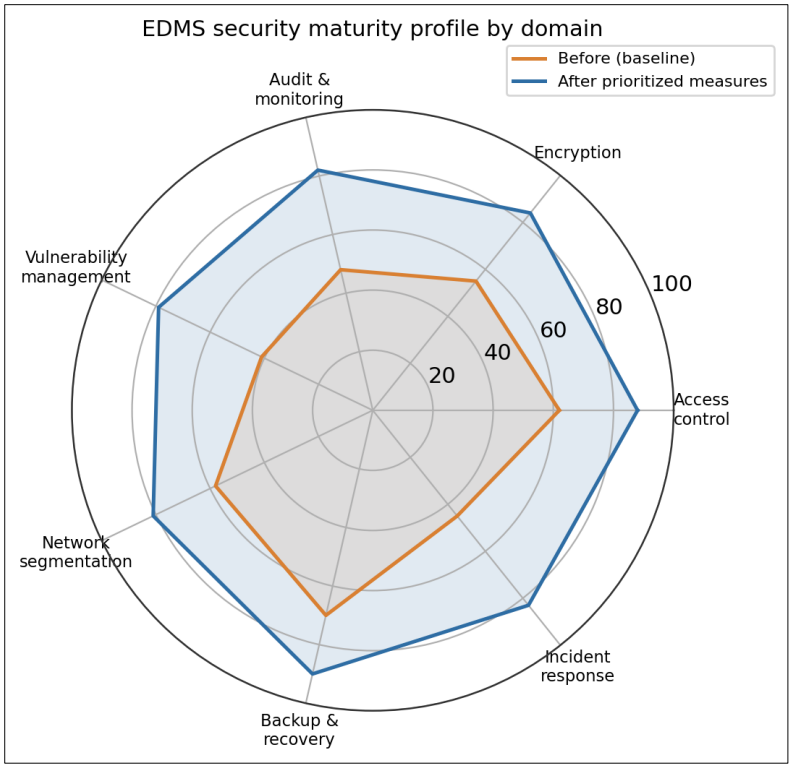

Figure 9. EDMS security-maturity profile by domain before and after prioritized measures

Decomposing the index gain by domain reveals the sources of improvement. The largest contribution came from closing gaps in vulnerability management, where the normalized domain risk decreased most substantially owing to the reduced age of unpatched vulnerabilities. Notable improvements were provided by strengthening auditing and developing response processes. The encryption and backup domains, initially close to the target level, changed only slightly, which confirms the selectivity of the recommendations: the system concentrates effort where the marginal return on invested resources is greatest.

4.6. Probability calibration and computational performance

Since risk management requires not only the ability to discriminate classes but also correct absolute probability values, the calibration of predictions was assessed. The reliability diagrams of the hybrid model are close to the diagonal, and the expected calibration error is noticeably lower than that of single gradient boosting, which tends to be overconfident. The improved calibration is explained by the participation of the Bayesian subsystem and the stacking meta-model. Good calibration is essential for the CSAI index, since domain risks are aggregated precisely from probabilistic estimates.

Computational performance confirms the practical applicability of the system. In the test environment, the full assessment of a single configuration state - fuzzification, ensemble inference, Bayesian-network update, and Fuzzy-AHP aggregation – is performed in tens of milliseconds, and batch processing of thousands of states scales linearly. The achieved performance is sufficient both for periodic auditing and for a near-continuous monitoring regime.

5. Comparative Analysis

This section consolidates the comparison of the proposed system with baseline and state-of-the-art methods along three axes - predictive quality together with computational and calibration cost, qualitative capabilities, and positioning relative to prior hybrid approaches - so that the claimed novelty is substantiated rather than declarative.

5.1. Quantitative comparison with state-of-the-art methods

Beyond predictive accuracy (table 8), a fair comparison must account for computational cost, inference latency, calibration quality, and interpretability, since these jointly determine fitness for continuous risk monitoring. Table 10 reports a comprehensive comparison in which the expected calibration error (ECE) and qualitative interpretability are added to the accuracy metrics. The modern boosting methods (XGBoost, LightGBM, CatBoost) occupy an intermediate position between the random forest and the stacking ensemble, confirming that they are strong but insufficient on their own for this task. The proposed hybrid attains the best ECE (0.021), reflecting well-calibrated probabilities, and retains high interpretability despite its compositional structure; its additional inference latency (about 12 ms, dominated by Bayesian inference) remains acceptable for both periodic and near-real-time use.

Table 10. Comprehensive comparison with baseline and state-of-the-art models

Method	F1	MCC	AUC	Train, s	Infer, ms	ECE	Interpretability
Logistic regression	0.842	0.805	0.864	3	0.4	0.061	High
Fuzzy (Mamdani)	0.881	0.851	0.901	-	0.9	0.052	High
ANFIS	0.899	0.869	0.918	95	1.2	0.048	Medium
Random forest	0.903	0.879	0.928	22	1.8	0.044	Medium
XGBoost	0.915	0.893	0.940	38	1.1	0.039	Medium
LightGBM	0.917	0.896	0.943	14	0.9	0.038	Medium
CatBoost	0.919	0.898	0.945	46	1.3	0.036	Medium
Stacking ensemble	0.922	0.902	0.951	120	3.5	0.034	Low
Hybrid (proposed)	0.946	0.927	0.972	142	12.0	0.021	High

5.2. Qualitative capability comparison

A comparison by qualitative capabilities (table 11) shows that, unlike isolated fuzzy, learnable, or Bayesian methods, the proposed system simultaneously provides learnability, causality, interpretability, domain contextualization, and a

single index. Checklist-based compliance approaches are interpretable but neither learnable nor causal; pure machine learning is learnable but opaque and context-agnostic; Bayesian networks are causal but do not learn from data without integration.

Table 11. Comparison of approaches by qualitative capabilities

Approach	Learnability	Causality	Interpretability	Domain context	Index
Compliance checklists	No	No	High	Partial	No
Fuzzy systems	No	No	High	Partial	Partial
Machine learning	Yes	No	Low	No	No
Bayesian networks	Partial	Yes	Medium	Partial	No
Proposed system	Yes	Yes	High	Yes	Yes

5.3. Comparison with prior hybrid approaches

To clarify the novelty, table 12 contrasts the proposed system with representative classes of hybrid solutions reported in the literature. Most published hybrids implement a sequential or two-component composition - fuzzy pre-processing followed by classification [7], [10], or Bayesian refinement of model outputs combined with scoring [12], [13] - and rarely provide consistent domain-level aggregation, adaptive weighting, or end-to-end explainability. Fuzzy-AHP multi-criteria schemes [20], [21] aggregate weights but lack a learnable component and a domain ontology, while machine-learning pipelines with post-hoc explanation [22], [23] add interpretability only as an afterthought. The proposed architecture differs along four dimensions: it preserves the autonomous outputs of three paradigms and reconciles them at the domain-aggregation level, which explains the super-additive gain observed in the ablation (table 10); it introduces a document-centric ontology of EDMS assets and threats with CIA weights; it replaces heuristic additive scales with a CSAI index whose weights are obtained by Fuzzy-AHP and grounded in multi-attribute utility theory [28], [29]; and it embeds explainability end-to-end rather than as an optional addition.

Table 12. Comparison with representative hybrid approaches

Representative class	Combined methods	Aggregation	Domain ontology	Adaptive weights	Explainability
Fuzzy + ML (sequential) [7], [10]	Fuzzy → ML	Heuristic	No	No	Partial
Bayesian + scoring [12], [13]	BN + CVSS	Scenario prob.	Partial	No	Partial
Fuzzy-AHP MCDM [20], [21]	Fuzzy-AHP	Weighted sum	Partial	No	Low
ML + post-hoc XAI [22], [23]	ML + SHAP	None	No	No	Post-hoc
Proposed system	Fuzzy + ML + BN	Fuzzy-AHP + CSAI	Yes	Yes	End-to-end

5.4. Statistical significance of the differences

The observed advantage of the hybrid model is statistically significant. A comparison of the per-fold F1 distributions using the Wilcoxon signed-rank test shows that the superiority over the stacking ensemble is stable at a significance level below 0.05, and over the single baseline models below 0.01. The approximate 95% confidence intervals reported in table 9 do not overlap with those of the single models for the F1 score and the Matthews coefficient, which further supports the stability of the observed performance differences. The narrowness of the intervals additionally indicates the stability of the estimates.

6. Discussion

The results confirm that integrating fuzzy inference, ensemble learning, and Bayesian modelling within a single architecture with a domain ontology yields a more accurate and stable assessment than any of the methods individually [6], [7], [8], [9], [11], [12]: the components are complementary, and the observed quality gain is super-additive. The CSAI index acts as a managerial interface, and the explainability layer links the estimate to the contribution of factors

and fired rules [22], [23]. In practice, the system is designed as an overlay on the existing infrastructure and produces prioritized recommendations ranked by expected risk reduction [3], [15], while the dynamics of the CSAI index serve as a measurable indicator of protection effectiveness.

For a balanced assessment, these benefits must be weighed against the operational costs and challenges of deployment. Initial setup is expertise-intensive: it requires adapting the ontology to the target environment, calibrating the membership functions [6], [7], and eliciting the conditional probability tables [11], [12], which is non-trivial for organizations without in-house security specialists. Ongoing maintenance adds further cost: the ensemble must be periodically re-trained as configurations and threats evolve, the Bayesian tables must be re-estimated as incidents accumulate [11], [12], and the Fuzzy-AHP weights must be revisited [20], [21] so that the model does not drift from the current environment. Because the labels and priors depend on expert input, sustained - if intermittent - expert involvement is needed, and a model-update procedure with version control and change auditing is advisable to keep assessments traceable. Integration with existing SIEM and monitoring pipelines, and the latency budget for near-real-time scoring, are additional practical considerations. These are manageable but real costs, and adopting organizations are recommended to plan explicitly for periodic re-calibration and for the human effort it entails.

7. Limitations of the study

The results should be interpreted with several limitations in mind. Sample limitations: although the dataset was collected from an operational electronic document management system and anonymized prior to analysis, it originates from a single institutional deployment. Consequently, the findings may not fully generalize to organizations with different architectures, operational practices, user populations, and threat landscapes. Additional validation on independent industrial and governmental EDMS deployments is therefore required. Ontology limitations: the domain ontology of assets and threats reflects a typical EDMS structure and, when transferred to systems with different architectures or regulatory contexts, may require adaptation of asset classes, threat categories, and CIA weights. Transfer limitations: the membership functions, Bayesian-network structure, and conditional probability tables were calibrated for the considered class of systems; therefore, deployment in other environments may require re-calibration and expert involvement during the initial setup stage. Threat-model limitations: the current implementation focuses on typical cybersecurity scenarios and does not explicitly model adaptive adversaries that intentionally manipulate observed features (adversarial behaviour). Finally, the expert-driven processes used for risk labelling and conditional probability elicitation introduce a degree of subjectivity, which, although mitigated through consistency checks and empirical calibration, cannot be completely eliminated. These limitations highlight the need for further independent validation before large-scale industrial deployment.

8. Conclusion

This paper proposed an intelligent system for the quantitative, reproducible, and explainable assessment of EDMS cybersecurity based on a hybrid six-layer architecture (Mamdani fuzzy inference, an ML stacking ensemble, and a Bayesian threat network with an attack graph), complemented by a domain ontology and the CSAI composite index with adaptive Fuzzy-AHP re-weighting. The model was experimentally shown to attain an F1 score of 0.946, a Matthews coefficient of 0.927, and an AUC of 0.972, outperforming single baseline models by 2.4 to 10.4 percentage points in F1; the contribution of each component was confirmed by ablation, and a comparative analysis demonstrated advantages over modern gradient-boosting methods and ANFIS in accuracy, calibration, and interpretability. The scientific novelty lies in the comprehensive integration of three paradigms with a domain ontology and a single explainable index for EDMS. Future work will address the automatic tuning of parameters, adversarial robustness, validation on industrial deployments, and real-time assessment.

More concretely, four directions are planned. First, adversarial-robustness experiments will test the system against adaptive adversaries that manipulate observable features, using evasion and data-poisoning attacks together with candidate defences such as adversarial training and robust feature selection. Second, a cross-organizational validation study will apply the framework to government, financial, and private-enterprise EDMS deployments to quantify external validity and the effort required for re-calibration. Third, a real-time streaming implementation will process configuration and telemetry events incrementally, maintaining a continuously updated CSAI rather than batch

snapshots. Fourth, the expert-driven components will be progressively automated, including data-driven learning of the conditional probability tables and semi-automatic extension of the ontology, to lower the setup and maintenance burden identified above.

9. Declarations

9.1. Author Contributions

Conceptualization: A.S., A.O., G.B., A.B., B.R., L.Z., M.S., A.N., and Z.L.; Methodology: A.O.; Software: A.S.; Validation: A.S., A.O., G.B., A.B., B.R., L.Z., M.S., A.N., and Z.L.; Formal Analysis: A.S., A.O., G.B., A.B., B.R., L.Z., M.S., A.N., and Z.L.; Investigation: A.S.; Resources: A.O.; Data Curation: A.O.; Writing Original Draft Preparation: A.S., A.O., G.B., A.B., B.R., L.Z., M.S., A.N., and Z.L.; Writing Review and Editing: A.O., A.S., G.B., A.B., B.R., L.Z., M.S., A.N., and Z.L.; Visualization: A.S.; All authors have read and agreed to the published version of the manuscript.

9.2. Data Availability Statement

The raw dataset cannot be made publicly available due to confidentiality requirements of the operational environment. Anonymized data samples may be made available upon reasonable request to the corresponding author, subject to institutional approval.

9.3. Funding

Not applicable.

9.4. Institutional Review Board Statement

Not applicable.

9.5. Informed Consent Statement

Not applicable.

9.6. Declaration of Competing Interest

The authors declare no conflicts of interest.

References

[1] W. Stallings and L. Brown, *Computer Security: Principles and Practice*, 4th ed. New York: Pearson, 2018. [Online]. Available: https://elibrary.pearson.de/book/99.150005/9781292220635

[2] W. Saffady, *Managing Electronic Records: Methods, Best Practices, and Technologies*, 5th ed. Lanham, MD: Rowman & Littlefield, 2017. [Online]. Available: https://search.worldcat.org/search?q=isbn:9781538104184

[3] National Institute of Standards and Technology, *Guide for Conducting Risk Assessments*, NIST Special Publication 800-30, Rev. 1, Sep. 2012, doi: 10.6028/NIST.SP.800-30r1.

[4] International Organization for Standardization, "Information security, cybersecurity and privacy protection – guidance on managing information security risks," ISO/IEC 27005:2022, 4th ed., Oct. 2022. [Online]. Available: https://www.iso.org/standard/80585.html

[5] A. Shameli-Sendi, R. Aghababaei-Barzegar, and M. Cheriet, "Taxonomy of information security risk assessment (ISRA)," *Computers & Security*, vol. 57, no. March, pp. 14–30, Mar. 2016, doi: 10.1016/j.cose.2015.11.001.

[6] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965, doi: 10.1016/S0019-9958(65)90241-X.

[7] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, vol. 7, no. 1, pp. 1–13, Jan. 1975, doi: 10.1016/S0020-7373(75)80002-2.

[8] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[9] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, vol. 2016, no. August, pp. 785–794, Aug. 13–17, 2016, doi: 10.1145/2939672.2939785.

[10] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, Second quarter 2016, doi: 10.1109/COMST.2015.2494502.

[11] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann, 1988, doi: 10.1016/C2009-0-27609-4.

[12] M. Frigault, L. Wang, A. Singhal, and S. Jajodia, "Measuring network security using dynamic Bayesian network," in *Proc. 4th ACM Workshop on Quality of Protection*, Alexandria, VA, USA, vol. 2008, no. October, pp. 23–30, Oct. 27, 2008, doi: 10.1145/1456362.1456368.

[13] Forum of Incident Response and Security Teams (FIRST), "Common vulnerability scoring system v4.0: specification document," FIRST.org, Inc., 2023. [Online]. Available: https://www.first.org/cvss/specification-document

[14] P. Mell, K. Scarfone, and S. Romanosky, "A complete guide to the common vulnerability scoring system version 2.0," FIRST.org, Inc., Jun. 2007. [Online]. Available: https://www.first.org/cvss/v2/guide

[15] R. S. Ross, *Managing Information Security Risk: Organization, Mission, and Information System View*, NIST Special Publication 800-39, vol. 2011, no. March, pp. 1-88, Mar. 2011, doi: 10.6028/NIST.SP.800-39.

[16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org/

[17] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York: Springer, 2013, doi: 10.1007/978-1-4614-6849-3.

[18] O. Sheyner, J. Haines, S. Jha, R. Lippmann, and J. M. Wing, "Automated generation and analysis of attack graphs," in *Proc. IEEE Symposium on Security and Privacy*, Oakland, CA, USA, vol. 2002, no. May, pp. 273–284, May 12–15, doi: 10.1109/SECPRI.2002.1004377.

[19] I. Kotenko and A. Chechulin, "A cyber attack modeling and impact assessment framework," in *Proc. 5th Int. Conf. Cyber Conflict (CyCon)*, Tallinn, Estonia, vol. 2013, no. June, pp. 119–142, Jun. 4–7, 2013, [Online]. Available: https://dblp.org/rec/conf/cycon/KotenkoC13.html

[20] T. L. Saaty, *The Analytic Hierarchy Process*. New York: McGraw-Hill, 1980. [Online]. Available: https://search.worldcat.org/title/The-analytic-hierarchy-process-:-planning-priority-setting-resource allocation/oclc/5352839

[21] D.-Y. Chang, "Applications of the extent analysis method on fuzzy AHP," *European Journal of Operational Research*, vol. 95, no. 3, pp. 649–655, Dec. 1996, doi: 10.1016/0377-2217(95)00300-2.

[22] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, Long Beach, CA, USA, vol. 2017, no. December, pp. 4765–4774, Dec. 4–9, 2017, doi: 10.48550/arXiv.1705.07874.

[23] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv:1702.08608, vol. 2017, no. February, pp. 1-13, Feb. 2017. doi: 10.48550/arXiv.1702.08608.

[24] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665–685, May 1993, doi: 10.1109/21.256541.

[25] S. Kabir and Y. Papadopoulos, "Applications of Bayesian networks and Petri nets in safety, reliability, and risk assessments: a review," *Safety Science*, vol. 115, no. July, pp. 154–175, Jul. 2019, doi: 10.1016/j.ssci.2019.02.009.

[26] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, Feb. 1992, doi: 10.1016/S0893-6080(05)80023-1.

[27] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. June, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[28] R. L. Keeney and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge, England: Cambridge University Press, 1993, doi: 10.1017/CBO9780511611308.

[29] E. Triantaphyllou, *Multi-Criteria Decision Making Methods: A Comparative Study*. New York: Springer, 2000, doi: 10.1007/978-1-4757-3157-6.

[30] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochimica et Biophysica Acta (BBA) – Protein Structure*, vol. 405, no. 2, pp. 442–451, Oct. 1975, doi: 10.1016/0005-2795(75)90109-9.

[31] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.

ISSN 2723-6471

[32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research*, vol. 12, no. October, pp. 2825–2830, Oct. 2011, doi: 10.48550/arXiv.1201.0490.

[33] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: a highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, Long Beach, CA, USA, 2017, pp. 3146–3154.

[34] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," in *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, Montréal, Canada, vol. 2018, no. December, pp. 6639–6649, Dec. 3–8, 2018,