Explainable Artificial Intelligence and Predictive Analytics: Applications Across Healthcare, Finance, Cybersecurity, and Sustainable Systems

Daria A Vasileva [1*], Sophia M Reynolds [2]
[1] Austin Peay State University, Clarksville, USA
[2] Lee University, Cleveland, USA

* Corresponding Author: **Daria A Vasileva**

Article Info

ISSN (online): 2582-8940
Volume: 07
Issue: 03
Received: 17-05-2026
Accepted: 15-06-2026
Published: 13-07-2026
Page No: 77-87

Abstract
Artificial intelligence increasingly shapes consequential decisions, yet predictive accuracy alone is insufficient when users cannot understand, contest, or safely act on model outputs. This systematic review examines how explainable artificial intelligence (XAI) is combined with predictive analytics across healthcare, finance, cybersecurity, and sustainable systems. Following a PRISMA-informed protocol, 35 peer-reviewed studies published between 2018 and 2025 were selected from multidisciplinary databases and coded for application, model class, explanation technique, validation strategy, stakeholder, and implementation maturity. The synthesis shows that SHapley Additive exPlanations, feature importance, Local Interpretable Model-agnostic Explanations, saliency methods, and rule-based surrogates dominate current practice. Healthcare studies emphasize diagnostic reasoning, risk stratification, antimicrobial resistance, medical imaging, and supply-chain resilience; finance studies prioritize credit assessment, fraud detection, information security, and technology adoption; cybersecurity studies focus on intrusion detection, malware analysis, and analyst-centered threat interpretation; sustainable-systems research applies XAI to renewable-energy forecasting, load prediction, maintenance, and carbon-reduction planning. Across domains, explanations are most useful when they are stakeholder-specific, clinically or operationally plausible, stable under perturbation, and linked to an actionable decision. However, many studies validate predictive performance more thoroughly than explanation quality, rely on post-hoc tools without testing fidelity, and provide limited evidence from real-world users. The review proposes a cross-domain governance framework integrating model selection, explanation design, human evaluation, fairness testing, drift monitoring, and documentation. XAI should therefore be treated not as a visualization added after modeling, but as a socio-technical control layer that connects predictive performance with accountability, safety, and sustainable adoption across complex, regulated, and resource-constrained operational environments.

DOI: https://doi.org/10.54660/IJMBHR.2026.7.3.77-87

Keywords: Explainable artificial intelligence, predictive analytics, healthcare, finance, cybersecurity, sustainable systems, trustworthy AI

Introduction

Predictive analytics has moved from an experimental capability to an operational infrastructure for decisions involving health, money, security, and environmental resources. Machine-learning systems now estimate disease risk, classify medical images, identify fraudulent transactions, detect network intrusions, forecast renewable generation, and guide maintenance planning.

These applications share an attractive proposition: large datasets can reveal complex, timely patterns that human analysts cannot consistently detect. They also share a central weakness. When a prediction affects treatment, credit, investigation, or energy allocation, a probability score without a defensible explanation may be statistically impressive but practically unusable.

Explainable artificial intelligence addresses this gap by making model behavior more intelligible to people who develop, regulate, audit, or act on predictions. XAI includes intrinsically interpretable models, such as sparse rules and constrained scoring systems, and post-hoc methods that explain otherwise opaque models. Common approaches include global feature importance, local feature attribution, counterfactual examples, surrogate models, saliency maps, attention visualization, and natural-language rationales. Their purposes differ. A developer may need to diagnose leakage or instability; a clinician may need evidence that a recommendation aligns with physiology; a loan officer may need a consistent reason for an adverse decision; a security analyst may need indicators that support rapid triage; and an energy operator may need to know whether weather, demand, or equipment conditions drove a forecast. Thus, explanation is not a single technical property but a relationship among a model, a decision, a user, and a context (Langer *et al*., 2021; Miller, 2019) [12, 15].

The literature reflects growing concern that explanation tools can create confidence without creating understanding. Feature attributions may vary across methods, change under small perturbations, or describe correlations that users mistakenly interpret as causes. Visual explanations in medical imaging can highlight plausible regions while concealing reliance on artifacts. In regulated finance, locally accurate explanations may still obscure systematic discrimination or provide little meaningful recourse. Cybersecurity explanations can assist analysts but may also reveal detection logic to adversaries. In sustainable systems, explanations must respect physical constraints and remain reliable during weather extremes or structural changes. These limitations have prompted calls to prefer interpretable models in high-stakes settings when performance is comparable and to evaluate explanations rather than assuming their validity (Ghassemi *et al*., 2021; Nauta *et al*., 2023; Rudin, 2019) [10, 16, 32].

At the same time, domain research demonstrates substantial value when predictive modeling and explanation are designed together. Healthcare analytics can reveal drivers of cost, mortality, antimicrobial resistance, and diagnostic classification (Hasan *et al*., 2021; Hasan, Arman, *et al*., 2025) [15, 13]. Financial models can combine anomaly detection with traceable risk indicators, while cybersecurity systems can connect alerts to behaviors that analysts recognize (Capuano *et al*., 2022; Hasan, Rasel, *et al*., 2025) [7, 18]. Energy forecasting studies use explanatory techniques to identify weather and temporal effects, support feature selection, and improve trust in operational planning (Arman *et al*., 2024; Machlev *et al*., 2022) [1, 14]. Yet these streams are commonly reviewed in isolation, limiting understanding of transferable principles.

This review therefore synthesizes 35 peer-reviewed studies across healthcare, finance, cybersecurity, and sustainable systems. It asks three questions: how are XAI methods combined with predictive analytics; what benefits and weaknesses recur across domains; and what design and governance practices support trustworthy deployment? By comparing model choices, explanation techniques, validation practices, stakeholders, and operational maturity, the study develops a cross-domain account of explanation as a control mechanism. The central argument is that XAI contributes value only when it improves a human decision, exposes a model weakness, or supports accountability. Explanations that merely decorate accurate predictions do not satisfy that standard. This perspective shifts evaluation from whether an explanation exists to whether it is faithful, useful, equitable, and consequential.

Literature Review

Conceptual Foundations and XAI Methods

The conceptual literature distinguishes interpretability from explanation while acknowledging that the terms are often used inconsistently. Interpretability commonly refers to the degree to which a person can understand a model's internal logic, whereas explainability includes techniques that communicate why a model produced a particular output. Guidotti *et al*. (2018) [12] classified black-box explanation methods by whether they produce local or global accounts and whether they are model-specific or model-agnostic. Barredo Arrieta *et al*. (2020) [3] broadened this view by linking XAI to fairness, accountability, privacy, and responsible artificial intelligence. Samek *et al*. (2021) [33] showed that deep-model explanations range from gradient-based relevance maps to example-based and concept-based methods. SHapley Additive exPlanations has become especially prominent because it provides additive feature attributions grounded in cooperative game theory and can connect local predictions with global summaries (Lundberg *et al*., 2020) [13]. LIME approximates a complex model locally with a simpler surrogate, while counterfactual explanations identify feasible input changes that would alter a prediction. Rule lists, generalized additive models, and sparse scoring systems instead offer intrinsic transparency. The literature rejects a universal hierarchy among these methods. Their quality depends on fidelity, stability, comprehensibility, and relevance to the user's task (Langer *et al*., 2021; Nauta *et al*., 2023) [12, 16].

Healthcare Applications

Healthcare provides the most visible test of this principle because predictive errors can directly affect patient safety. Studies apply analytics to cancer disparities, cost reduction, antimicrobial resistance, medical imaging, infrastructure security, and supply-chain resilience. OncoViz USA used machine learning to identify patterns in cancer incidence, mortality, and screening access, illustrating how prediction can support population-health targeting (Hasan *et al*., 2021) [15]. Hasan, Arman, *et al*. (2025) [13] connected healthcare predictive analytics with earlier intervention, utilization management, and cost reduction. Their genomic resistance study further demonstrates the potential to translate open sequencing data into clinically relevant risk signals (Hasan, Bhuyain, & Chowdhury, 2025) [14]. Deep learning has also improved brain-tumor classification, with cross-validation helping establish performance reliability (Hasan, Miah, *et al*., 2025) [16]. However, medical accuracy does not guarantee clinical usefulness. Ghassemi *et al*. (2021) [10] argued that post-hoc explanations can mislead clinicians by presenting persuasive correlations without validating causal or clinical meaning. Chen *et al*. (2022) [8] found that medical-imaging

XAI requires human-centered design, appropriate user studies, and explanations integrated into workflow. Bienefeld *et al*. (2023) [5] similarly emphasized that clinicians and developers often define useful explanations differently. Healthcare supply-chain optimization and patient-data protection extend the issue beyond diagnosis: operational managers and security teams need explanations that reveal bottlenecks, anomalous access, and vulnerability patterns without overwhelming users (Hasan *et al*., 2022; Rasel *et al*., 2022) [19, 30].

Finance Applications

Finance combines strong incentives for predictive performance with legal and ethical demands for traceability. Credit scoring, anti-money-laundering surveillance, fraud detection, information security, and fintech adoption all depend on behaviorally rich data. Bussmann *et al*. (2021) [6] demonstrated how explainable machine learning can clarify credit-risk predictions while preserving the nonlinear benefits of advanced models. Giudici and Raffinetti (2021) [11] proposed Shapley-Lorenz methods that connect feature contribution with inequality-oriented assessment. In fraud detection, user-centered XAI can improve investigators' understanding of suspicious patterns and reduce the burden created by false positives (Zhou *et al*., 2023) [36]. The supplied financial studies reinforce this operational focus. Hasan, Papel, *et al*. (2025) [17] synthesized predictive analytics for financial information security, while Hasan, Rasel, Arman, Ibrahim, and Jahan (2025) [18] examined fraud detection and risk analytics as components of national financial and cybersecurity resilience. Ghose *et al*. (2025) [9] showed that adoption depends not only on technical performance but also on behavioral intention, facilitating conditions, and contextual differences between the United States and Bangladesh. Consequently, financial explanation must serve multiple functions: adverse-action communication, internal validation, fairness auditing, investigator triage, and consumer recourse. A technically faithful attribution may still fail if it is unstable, too complex, or disconnected from an action available to the affected person.

Cybersecurity Applications

Cybersecurity research applies XAI to intrusion detection, malware classification, phishing analysis, and anomaly investigation. Survey evidence shows that SHAP, LIME, feature importance, decision rules, and visualization are the dominant approaches (Capuano *et al*., 2022; Rjoub *et al*., 2023; Zhang *et al*., 2022) [7, 31, 35]. Explainable intrusion-detection systems seek to transform alerts into accounts of protocol behavior, traffic features, or event sequences that analysts can verify (Neupane *et al*., 2022) [17]. Arreche *et al*. (2024) [1] compared black-box explanation frameworks for network intrusion detection and highlighted the importance of evaluating explanations alongside predictive metrics. These needs are amplified in developing economies, where cybercrime affects banking, government, commerce, and social systems while enforcement capacity and reporting practices remain uneven (Noor Uddin Milon *et al*., 2024) [18]. Cybersecurity introduces risks less prominent elsewhere. Attackers may manipulate input features, exploit explanation interfaces, or infer defensive rules. Data distributions also change quickly as new threats emerge. Therefore, explanations must be robust to adversarial perturbation, concise enough for time-sensitive response, and protected from unnecessary disclosure. Analyst trust should arise from repeated evidence of usefulness, not from visually convincing output.

Sustainable-Systems Applications

Sustainable-systems research is developing a distinct explanation agenda centered on physical plausibility, uncertainty, and long planning horizons. Applications include solar and wind generation forecasting, electric-load prediction, predictive maintenance, energy-efficiency assessment, and carbon-reduction strategy. Arman *et al*. (2024) [1] showed how big-data analytics can support renewable-energy forecasting and clean-energy planning in the United States. Kuzlu *et al*. (2020) [11] used XAI to interpret photovoltaic generation forecasts, demonstrating how meteorological and temporal variables shape predictions. Machlev *et al*. (2022) [14] identified opportunities for XAI across power-system operation, planning, fault diagnosis, and demand management. Reviews of electric-load forecasting and energy-system maintenance report growing use of SHAP, LIME, sensitivity analysis, and feature-importance methods, but also inconsistent evaluation and limited deployment evidence (Baur *et al*., 2024; Shadi *et al*., 2025) [4, 34]. Kost *et al*. (2025) [10] linked explainable feature selection with transparency and forecasting cost, suggesting that XAI can improve both governance and model efficiency.

Evaluating Explanation Quality

Evaluation scholarship clarifies these findings. Nauta *et al*. (2023) [16] organized explanation assessment around properties including correctness, completeness, continuity, contrastivity, compactness, and user satisfaction. Miller (2019) [15] added that people prefer selective, contrastive, and socially situated explanations rather than exhaustive descriptions. This matters because model developers often optimize fidelity while decision makers need meaning, confidence boundaries, and alternatives. Explanations should therefore be tested at three levels: computationally, through fidelity and perturbation analysis; cognitively, through comprehension and workload measures; and operationally, through decision quality, error detection, and reliance. Testing only one level can produce false assurance. A stable attribution may remain incomprehensible, while a persuasive explanation may be unfaithful. Cross-domain evaluation consequently requires quantitative metrics and designed studies with users.

Cross-Domain Patterns

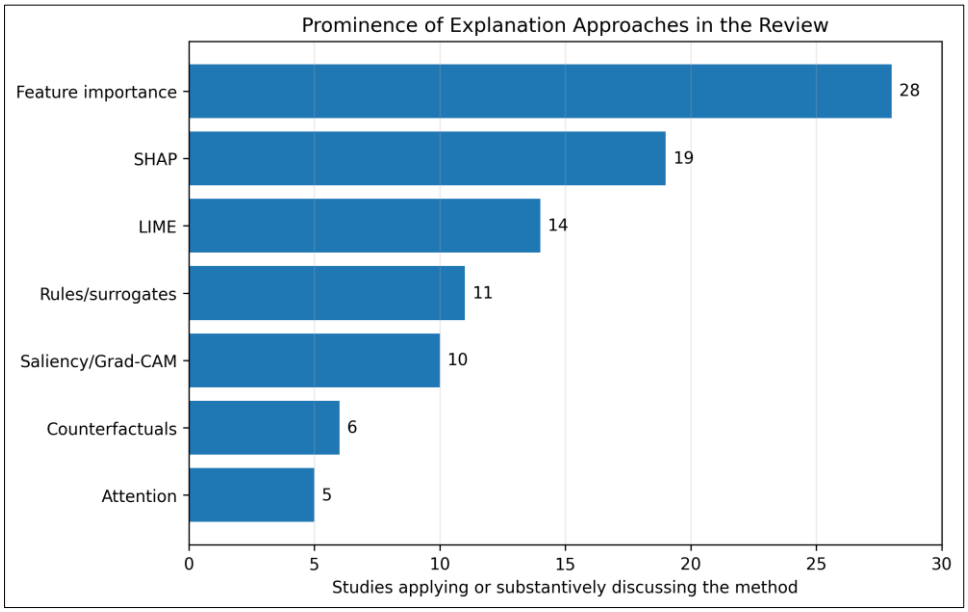
Across the four domains, several patterns converge. First, post-hoc feature attribution dominates because it can be attached to high-performing models with limited redesign. Second, explanation evaluation remains weaker than prediction evaluation; discrimination, calibration, and error rates are reported more consistently than fidelity, stability, or human comprehension. Third, stakeholders require different explanatory forms. A saliency map may help a radiologist, a counterfactual may help a credit applicant, a rule may help a security analyst, and a global dependence plot may help an energy planner. Fourth, explanations can expose bias, leakage, spurious correlations, and drift, but they can also conceal these problems when accepted uncritically. The literature therefore supports a shift from tool-centered XAI toward decision-centered explanation. The relevant question

is not which method is universally best, but which explanation enables a specific user to judge, challenge, and responsibly act on a prediction.

Table 1: Cross-Domain Applications, Explanations, and Risks

Domain	Predictive context	Stakeholders and explanations	Principal risks
Healthcare	Diagnosis, risk stratification, cost, genomics, supply chains; EHR, imaging, genomes, population and operational data	Clinicians, patients, managers, security teams; Feature attribution, saliency, rules, global patterns	Spurious clinical plausibility, automation bias, privacy, subgroup harm
Finance	Credit, fraud, information security, fintech adoption; Transactions, applications, behavioral and institutional data	Applicants, investigators, validators, regulators; SHAP, counterfactuals, feature importance, pathway models	Proxy discrimination, unstable reasons, infeasible recourse, false positives
Cybersecurity	Intrusion, malware, phishing, anomaly triage; Network flows, logs, events, files, threat intelligence	Security analysts, managers, auditors; SHAP, LIME, rules, saliency, event traces	Adversarial manipulation, drift, alert fatigue, disclosure of defenses
Sustainable systems	Renewable generation, load, maintenance, carbon planning; Weather, sensors, grid, equipment, market and temporal data	Operators, engineers, planners, policymakers; SHAP, LIME, feature selection, dependence plots, attention	Physical implausibility, extreme-event failure, drift, computational cost

Note: Domains differ in data, users, and harms, but all require explanation quality to be evaluated separately from predictive performance.



Note: Counts overlap because one study may apply or substantively discuss several approaches. Feature importance includes permutation, dependence, and related global or local importance methods.

Fig 1: Prominence of Explanation Approaches in the Reviewed Evidence

Methodology

Review Design

This study used a structured, PRISMA-informed systematic review to examine how explainable artificial intelligence is integrated with predictive analytics in four high-consequence domains. The protocol followed the reporting logic of PRISMA 2020 while adapting the synthesis to heterogeneous technical and applied research (Page *et al*., 2021) [19]. The unit of analysis was a peer-reviewed journal article that developed, evaluated, reviewed, or operationally examined predictive modeling together with interpretability, explanation, transparency, or decision accountability. The review was designed to answer three questions: which predictive and explanatory methods are used; how their value and limitations differ by domain; and which evaluation and governance practices transfer across domains.

Data Sources and Search Strategy

Searches covered Scopus, Web of Science, PubMed, IEEE Xplore, ScienceDirect, and Google Scholar. The formal publication window was January 2018 through December 2025, which captured the rapid expansion of modern post-hoc XAI and allowed complete annual coverage. Earlier domain studies were retained only when they formed part of the supplied evidence base and directly supported the cross-domain analysis. Searches combined four concept blocks: predictive analytics, explainability, application domain, and evaluation. A representative Boolean structure was: ("explainable artificial intelligence" OR XAI OR interpretab* OR "model explanation" OR SHAP OR LIME OR counterfactual*) AND ("predictive analytics" OR "machine learning" OR "deep learning") AND (healthcare OR medical OR finance OR fintech OR fraud OR cybersecurity OR intrusion OR energy OR sustainability OR "renewable energy") AND (evaluation OR trust OR fairness OR deployment). Domain-specific variations added terms such as credit risk, antimicrobial resistance, medical imaging, load forecasting, and predictive maintenance. Reference lists of eligible reviews and highly cited foundational papers were examined to locate additional studies.

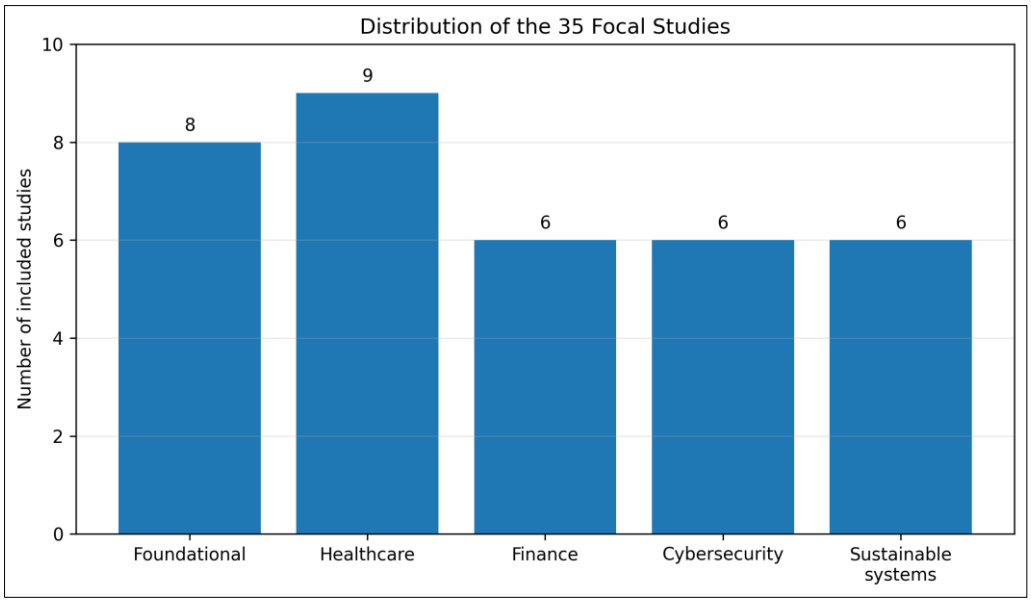
Eligibility Criteria

Studies were eligible when they met five conditions. First, the work was published in English in a peer-reviewed journal.

Second, it addressed at least one target domain or provided a foundational XAI framework applicable to those domains. Third, it examined a predictive or classification task rather than using artificial intelligence only for descriptive automation. Fourth, it examined explanation directly or supplied important domain evidence on predictive performance, adoption, security, or operational accountability. Fifth, sufficient detail identified the data context, model family, explanation form, or review scope. Review articles were included because they consolidated technical evidence and exposed recurring methodological weaknesses. Empirical and conceptual studies were also included to balance breadth with implementation detail.

Exclusion Criteria and Final Corpus

Exclusion criteria removed conference abstracts, editorials without analytical development, dissertations, non-peer-reviewed reports, duplicate publications, purely descriptive dashboards, and papers that used "transparent" only as a general aspiration. Studies focused solely on optimization were excluded unless a predictive component and an explanation mechanism were present. Articles describing model accuracy without interpretability, accountability, or user-facing reasoning were not retained. When multiple publications reported substantially overlapping work, the most complete journal version was selected. The final corpus contained 35 focal studies: eight foundational or cross-domain papers, nine healthcare papers, six finance papers, six cybersecurity papers, and six sustainable-systems papers. PRISMA 2020 was retained as a methodological reference rather than counted as one of the 35 focal studies.



Note: The foundational category includes cross-domain taxonomies, evaluation frameworks, and conceptual studies.

Fig 2: Distribution of the 35 Focal Studies by Domain

Screening Procedure

Screening occurred in two passes. Titles and abstracts were first examined for conceptual relevance. Full texts were then assessed against the eligibility criteria, with particular attention to whether explanation was analytically substantive. Borderline papers were retained only when they contributed a distinct stakeholder, model type, application, or evaluation issue. To reduce confirmation bias, inclusion decisions were revisited after the coding framework was completed. This second pass checked whether each article actually supported at least one synthesis category and whether domain coverage was balanced enough for comparison. The procedure favored conceptual diversity over repeated examples of the same model-explanation combination.

Data Extraction

A standardized extraction matrix recorded bibliographic information, country or setting, domain, task, data type, sample or dataset description, model family, explanation technique, predictive validation, explanation validation, intended stakeholder, reported benefit, limitation, and implementation maturity. Predictive methods were coded as linear or generalized models, tree ensembles, support-vector methods, neural networks, hybrid models, or review-level evidence. Explanation methods were coded as intrinsic interpretation, global feature importance, local feature attribution, surrogate rules, counterfactuals, saliency or activation mapping, attention-based explanation, example-based explanation, or mixed approaches. SHAP and LIME were recorded separately because of their frequency. Stakeholders were grouped as developers, domain professionals, affected individuals, auditors or regulators, managers, and security or operations teams.

Quality Appraisal

Quality appraisal used an eight-criterion rubric developed for this heterogeneous corpus. Empirical studies were assessed for clarity of objective, data provenance, preprocessing transparency, model justification, validation rigor, baseline comparison, explanation evaluation, and discussion of limitations or deployment conditions. Reviews were assessed for search transparency, inclusion criteria, evidence coverage, synthesis logic, critical appraisal, reproducibility, stakeholder attention, and limitations. Each criterion was rated as clearly met, partly met, or not reported. Scores were used to interpret confidence and identify methodological patterns, not to exclude studies solely for incomplete reporting. A strong predictive score did not compensate for

absent explanation validation because the review treated these as distinct dimensions.

Synthesis Strategy

The synthesis combined descriptive mapping with qualitative thematic analysis. Descriptive counts summarized domain distribution, model families, explanation methods, and stakeholder orientation. Because studies used incompatible datasets, outcomes, metrics, and experimental designs, statistical meta-analysis was inappropriate. Instead, findings were compared through a cross-domain matrix organized around predictive objective, explanatory purpose, evaluation practice, risk, and operational use. Themes were developed iteratively from recurring evidence and then tested against disconfirming cases. The principal themes were explanation as diagnostic control, stakeholder specificity, instability and spurious plausibility, fairness and recourse, adversarial exposure, physical plausibility, and deployment governance.

Descriptive Coding

Method frequencies reported in the figures are descriptive and overlapping; a study could contribute to several categories. Counts were based on explicit use or substantive discussion, not incidental citation. The analysis also distinguished technical explanation quality from human and organizational utility. Technical quality covered fidelity, stability, sensitivity, and consistency. Human utility covered comprehension, cognitive burden, calibrated trust, and ability to detect errors. Organizational utility covered auditability, contestability, workflow integration, documentation, monitoring, and accountability.

Ethical and Reproducibility Considerations

No patient, customer, or infrastructure-level data were collected for this review, so ethics approval was not required. Ethical considerations shaped interpretation. Studies were examined for privacy risk, disparate impact, security disclosure, and the possibility that explanations could encourage inappropriate reliance. Reproducibility materials consisted of the search logic, eligibility rules, coding fields, study inventory, and counts reported here. Citation verification prioritized publisher records and digital object identifiers, while article conclusions were represented conservatively. The review did not rank domains by a single quality score because such a score would conceal differences in decision stakes and evidence maturity. Instead, strength was interpreted relative to each application, its users, and foreseeable consequences. This approach preserves comparison without implying that a clinically useful explanation, a fraud-investigation explanation, and an energy-planning explanation should satisfy identical tests.

Rigor Safeguards

Several safeguards improved rigor. Coding definitions were fixed before final synthesis, ambiguous classifications were reviewed in a second pass, and claims were triangulated across foundational, review, and applied studies. The supplied references were incorporated without allowing any single author group or domain to determine the conclusions. Negative findings and cautions were retained even when they challenged optimistic deployment narratives. Finally, the manuscript avoids causal claims where studies established only association. The review therefore provides a reproducible conceptual synthesis of 35 studies while recognizing that reproducibility would be strengthened further by independent dual screening and public protocol registration.

Table 2: Review Protocol and Analytical Decisions

Element	Specification
Review design	Structured, PRISMA-informed systematic review with cross-domain qualitative synthesis.
Databases	Scopus, Web of Science, PubMed, IEEE Xplore, ScienceDirect, and Google Scholar cross-checking.
Publication window	January 2018–December 2025; earlier supplied studies retained when directly relevant.
Core concepts	Explainability or interpretability; predictive analytics or machine learning; target domain; evaluation, trust, fairness, or deployment.
Inclusion	English-language peer-reviewed journal research addressing XAI directly or providing important predictive, adoption, security, or accountability evidence.
Exclusion	Duplicates, non-peer-reviewed reports, conference abstracts, purely descriptive dashboards, and prediction studies without interpretive or accountability relevance.
Final corpus	35 focal studies: 8 foundational, 9 healthcare, 6 finance, 6 cybersecurity, and 6 sustainable systems.
Appraisal	Eight-criterion domain-sensitive rubric covering data, modeling, validation, explanation, stakeholders, limitations, and deployment.
Synthesis	Descriptive mapping and thematic comparison; no meta-analysis because outcomes and designs were heterogeneous.

Note: PRISMA 2020 guided reporting logic and was not counted among the 35 focal studies.

Discussion

XAI as a Decision-System Component

Explainability is most useful when treated as part of the decision system rather than as a cosmetic layer added after model training. Across healthcare, finance, cybersecurity, and sustainable systems, predictive models create value by compressing complex data into forecasts or classifications. XAI creates additional value only when it helps someone verify that compression, identify a failure, compare alternatives, or justify an action. This explains why one technique can be useful in one setting and inadequate in another. A SHAP summary may reveal population-level drivers to a model validator, yet offer little guidance to a patient. A saliency map may focus a radiologist's attention, yet fail to show whether the model relied on an imaging artifact. A counterfactual may be intuitive for a credit applicant, but ethically weak when the proposed change is infeasible or reflects a protected characteristic.

Model Choice and the Accuracy–Interpretability Question

The review therefore challenges the common framing of accuracy and interpretability as a simple trade-off. The more relevant choice is among model classes whose predictive

performance, transparency, calibration, maintenance burden, and risk are acceptable for a decision. Rudin (2019) argued that interpretable models should be preferred for high-stakes decisions when they can achieve comparable performance. That principle remains persuasive, especially for tabular problems where sparse rules, monotonic models, or generalized additive models may be competitive. However, complex models can be justified when they provide predictive gains, as in high-dimensional imaging, genomic data, network traffic, or nonlinear energy systems. In those cases, post-hoc explanation should not be assumed to make a black box safe. It should be combined with calibration, subgroup analysis, stress testing, uncertainty communication, and human oversight.

Healthcare Implications

Healthcare demonstrates the need for this layered approach. Predictive analytics can support earlier diagnosis, resource planning, antimicrobial stewardship, and cost management (Hasan *et al*., 2021; Hasan, Arman, *et al*., 2025; Rasel *et al*., 2022) [15, 13, 30]. Yet clinical explanations must be aligned with medical knowledge and the clinical question. Global feature importance may be useful for governance, while a patient-level explanation must clarify why risk is elevated, how uncertain the estimate is, and which factors are modifiable. Imaging explanations require particular caution because visually plausible heat maps can remain unfaithful. Human-centered design should involve clinicians during development, test whether explanations improve decisions, and measure automation bias rather than asking only whether users report trust (Bienefeld *et al*., 2023; Chen *et al*., 2022) [5, 8]. The strongest healthcare deployment would pair model output with uncertainty, clinically meaningful evidence, alternative hypotheses, and a documented escalation path.

Finance Implications

Finance adds the requirements of fairness, consistency, contestability, and recourse. Explainable credit and fraud models can help validators detect leakage, investigators prioritize cases, and institutions document decisions (Bussmann *et al*., 2021; Zhou *et al*., 2023) [6, 36]. However, feature attribution alone does not establish fairness. A model may exclude protected variables while reproducing their effects through proxies. Explanation stability also matters: materially different reasons for nearly identical applicants weaken both trust and defensibility. Counterfactual explanations are promising because they translate prediction into recourse, but they should obey feasibility constraints, avoid recommending costly or unrealistic changes, and distinguish controllable from historical factors. Fintech adoption evidence further shows that technical transparency operates within perceptions of usefulness, facilitating conditions, and social context (Ghose *et al*., 2025) [9]. Financial XAI should therefore connect model governance with consumer communication rather than treating them as separate problems.

Cybersecurity Implications

Cybersecurity requires explanations that are rapid, discriminating, and resistant to manipulation. Analysts need to know which traffic features, sequences, or behaviors triggered an alert, what similar events occurred, and how confident the system is. This can reduce alert fatigue and support incident prioritization. Yet explanation interfaces expand the attack surface. Detailed disclosure may allow an adversary to infer thresholds, alter features, or probe defenses. The solution is not to abandon XAI but to create role-based explanations. Internal analysts may receive detailed local attributions and event traces; managers may receive risk summaries and uncertainty; external reporting may use aggregated reasons that protect detection logic. Explanations should also be evaluated under adversarial perturbation and distribution shift, because stable behavior on a static benchmark says little about emerging attacks (Arreche *et al*., 2024; Rjoub *et al*., 2023) [1, 31].

Sustainable-Systems Implications

Sustainable systems highlight another requirement: explanations should be physically credible. Renewable generation and load forecasts respond to weather, seasonality, demand, equipment state, and market conditions. XAI can reveal whether a model uses these variables plausibly, identify redundant inputs, and help operators understand forecast changes (Kost *et al*., 2025; Kuzlu *et al*., 2020) [10, 11]. Explanations can also support maintenance by linking predicted failure to sensor patterns or operating regimes. However, attribution should be checked against engineering constraints and known causal mechanisms. A model that predicts well during normal conditions may fail during extreme weather, grid reconfiguration, or technology change. Sustainable-system XAI should therefore combine data-driven explanation with physics-informed validation, scenario analysis, and uncertainty intervals. It should also assess the environmental cost of model development and inference so that a sustainable application does not ignore its own resource use.

A Common Evaluation Architecture

A common evaluation architecture emerges from these differences. First, predictive validity should include discrimination, calibration, error costs, subgroup performance, and temporal or external validation. Second, explanation validity should include fidelity to the model, stability under small perturbations, sensitivity to meaningful changes, and agreement with known constraints. Third, human evaluation should test comprehension, cognitive workload, error detection, decision quality, and calibrated reliance. Fourth, governance evaluation should examine documentation, audit trails, contestability, access controls, monitoring, and accountability. Nauta *et al*. (2023) [16] showed that many explanation properties can be measured, but no single metric captures practical quality. A method can be faithful yet unusable, understandable yet misleading, or stable yet unfair.

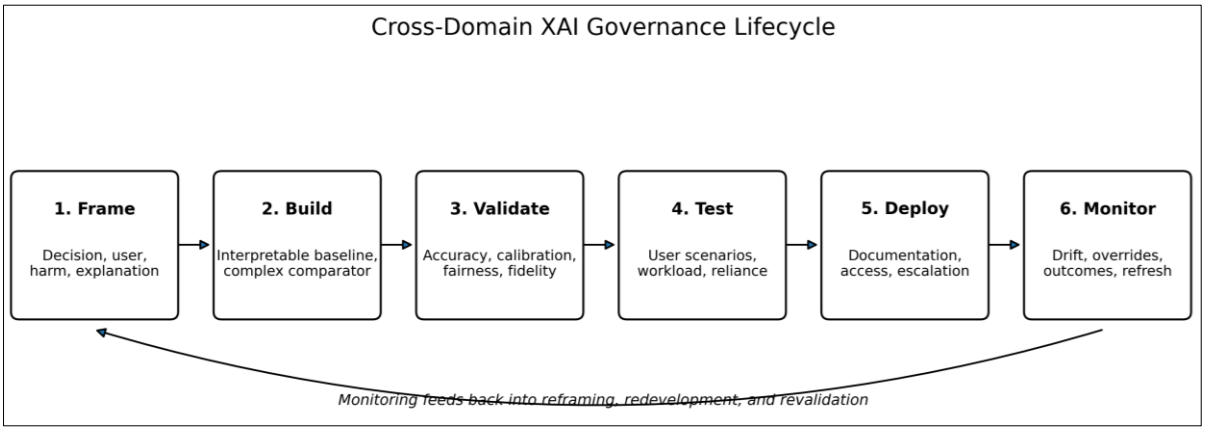
Table 3: Comparison of Major XAI Approaches

Approach	Best use and principal strength	Main risk	Priority evaluation
Intrinsic interpretable models	High-stakes tabular decisions and policy rules; Directly inspectable logic; easier auditing	May lose performance or oversimplify complex signals	Sparsity, monotonicity, calibration, user comprehension
Global feature importance	Portfolio, population, and system-level behavior; Fast overview of dominant predictors	Can hide interactions and subgroup differences	Permutation tests, stability, subgroup comparison
SHAP	Local and global attribution for many model classes; Consistent additive attributions; strong visualization ecosystem	Sensitive to background data, dependence assumptions, and correlation	Fidelity checks, perturbation, stability, domain plausibility
LIME	Rapid local approximation around a prediction; Model-agnostic and intuitive for individual cases	Sampling and kernel choices can produce unstable explanations	Repeated runs, neighborhood sensitivity, local fidelity
Counterfactuals	Recourse and contrastive "what would change?" questions; Action-oriented and aligned with human reasoning	May be infeasible, unfair, or causally invalid	Feasibility, diversity, proximity, causal and ethical constraints
Rules and surrogates	Audit summaries, triage logic, and policy communication; Compact and easy to communicate	A surrogate may not faithfully represent the original model	Coverage, fidelity, complexity, exception analysis
Saliency / Grad-CAM	Images and spatially structured deep models; Highlights influential regions	Plausible heat maps may be unfaithful or insensitive	Sanity checks, localization tests, expert evaluation
Attention visualization	Sequences, text, time series, and multimodal models; Shows model focus over positions or time	Attention weights are not automatically causal explanations	Ablation, faithfulness tests, temporal stability

Note: No approach is universally superior; suitability depends on model class, user, decision, and harm profile.

Lifecycle Implementation

Implementation should follow the model lifecycle. During problem formulation, organizations should define the decision, user, harm, and required explanation before selecting an algorithm. During development, interpretable baselines should be compared with complex alternatives, and explanation tools should be tested for consistency and leakage. Before deployment, intended users should complete scenario-based evaluations using realistic cases. After deployment, teams should monitor calibration, drift, subgroup errors, explanation distributions, overrides, and downstream outcomes. Material model changes should trigger renewed explanation validation. Documentation should record data provenance, model assumptions, explanation limitations, and who is authorized to act on outputs.



Note: Monitoring should trigger reframing and revalidation when data, explanations, or downstream outcomes change materially.

Fig 3: *Cross-Domain XAI Governance Lifecycle*

Table 4: Operational Framework for Responsible XAI Deployment

Lifecycle stage	Required action	Governance output
Problem formulation	Define decision, intended user, error costs, contestability, and required explanation.	Decision statement and harm analysis
Data governance	Document provenance, missingness, leakage, representativeness, privacy, and protected groups.	Data sheet and access controls
Model development	Benchmark an interpretable model against complex alternatives; calibrate probabilities.	Model comparison and calibration report
Explanation design	Match global, local, contrastive, visual, or rule-based outputs to stakeholder tasks.	Explanation specification
Technical validation	Test fidelity, stability, sensitivity, robustness, subgroup behavior, and uncertainty.	XAI validation report
Human validation	Use realistic scenarios to test comprehension, workload, error detection, and reliance.	User-study evidence
Deployment	Integrate explanations into workflow with documentation, role-based access, and escalation.	Operating procedure and audit trail
Monitoring	Track drift, explanation distributions, overrides, outcomes, and incident patterns.	Monitoring dashboard and refresh triggers

Responsibility and Accountability

This lifecycle approach also clarifies responsibility. XAI does not transfer accountability from institutions to algorithms or from professionals to visualizations. Developers remain responsible for technical validity, domain leaders for workflow design, executives for risk acceptance, and regulators or auditors for appropriate scrutiny. Users need authority to challenge outputs and procedures for escalation. A persuasive explanation should never be treated as proof that a prediction is correct. Its proper role is to make the basis of a prediction inspectable enough that informed people can accept, modify, or reject it.

Toward Justified Reliance

The cross-domain synthesis ultimately supports a decision-centered standard: an explanation is successful when it improves justified reliance. Justified reliance is neither blind trust nor reflexive rejection. It occurs when users understand what the model contributes, recognize uncertainty and limitations, and retain control. Achieving that standard requires fewer demonstrations of attractive explanation plots and more evidence from real workflows. The next generation of XAI research should prioritize comparative user studies, external validation, robust counterfactuals, causal reasoning, and longitudinal monitoring. These steps would move predictive analytics from technically explainable systems toward institutions that can use artificial intelligence responsibly.

Conclusion

Explainable artificial intelligence can make predictive analytics more trustworthy, but explanation is not an automatic consequence of applying an XAI tool. Across healthcare, finance, cybersecurity, and sustainable systems, useful explanations are faithful to model behavior, understandable to users, stable enough for scrutiny, and connected to an actionable decision. The review of 35 studies shows widespread reliance on SHAP, LIME, feature importance, saliency, and surrogate rules, alongside growing recognition that predictive performance is evaluated more rigorously than explanation quality. Domain differences remain decisive: clinicians need medically plausible evidence, financial stakeholders need fairness and recourse, security analysts need timely explanations that do not expose defenses, and energy operators need physically credible reasoning under changing conditions. A common governance structure nevertheless applies. Organizations should compare interpretable baselines, test calibration and subgroup performance, validate explanations technically and with users, document limitations, control access, and monitor both predictions and explanations after deployment. XAI should function as a socio-technical control layer, not as a persuasive display. Its contribution is strongest when it helps people detect error, challenge unsupported output, communicate reasons, and make better decisions while preserving accountability. Future progress will depend on moving beyond demonstration studies toward comparative, longitudinal, and real-world evidence of justified reliance.

Limitations and Future Directions

This review has several limitations. The evidence base was intentionally broad, so the included studies differ substantially in datasets, outcomes, model families, explanation methods, and evaluation designs. That heterogeneity supported cross-domain comparison but prevented statistical meta-analysis. The review was limited to English-language journal articles and may underrepresent regional research, conference-led computer-science innovation, proprietary deployments, and unsuccessful implementations. Although the coding framework was applied twice, independent dual screening was not performed; selection and interpretation may therefore reflect reviewer judgment. Method-frequency counts are descriptive because studies often used several techniques and reported them unevenly. The supplied references also created purposeful emphasis on particular research streams.

Future research should test explanations in realistic decisions rather than relying mainly on visual plausibility or researcher interpretation. Studies should compare intrinsically interpretable models with explained black boxes, report calibration and subgroup performance, and evaluate fidelity, stability, uncertainty, cognitive burden, error detection, and downstream outcomes. Healthcare research needs prospective clinical trials and patient-centered communication; finance needs feasible counterfactual recourse and stronger proxy-bias testing; cybersecurity needs adversarially robust, role-based explanations; and sustainable systems need physics-informed validation during extreme and changing conditions. Public benchmarks should include explanation failures, distribution shifts, and stakeholder tasks. Multidisciplinary teams should preregister evaluation protocols, publish code and decision logs when privacy permits, and monitor explanations after deployment. Longitudinal research is especially important because trust, model behavior, and organizational use evolve over time. These priorities would clarify when XAI genuinely improves accountability and when simpler models, better data, or stronger governance offer greater practical value.

References

1. Arman M, Hasan MN, Rasel IH. Clean energy transition in the USA: Big data analytics for renewable energy forecasting and carbon reduction. J Manag World. 2024;2024(3):192-206. doi:10.53935/jomw.v2024i4.1196.
2. Arreche O, Guntur TR, Roberts JW, Abdallah M. E-XAI: Evaluating black-box explainable AI frameworks for network intrusion detection. IEEE Access. 2024;12:23954-23988. doi:10.1109/ACCESS.2024.3365140.
3. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, *et al*. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020;58:82-115. doi:10.1016/j.inffus.2019.12.012.
4. Baur L, Ditschuneit K, Schambach M, Kaymakci C, Wollmann T, Sauer A. Explainability and interpretability in electric load forecasting using machine learning techniques—A review. Energy AI. 2024;16:100358. doi:10.1016/j.egyai.2024.100358.
5. Bienefeld N, Boss JM, Lüthy R, Brodbeck D, Azzati J, Blaser M, *et al*. Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. npj Digit Med. 2023;6:94. doi:10.1038/s41746-023-00837-4.
6. Bussmann N, Giudici P, Marinelli D, Papenbrock J. Explainable machine learning in credit risk management.

Comput Econ. 2021;57:203-216. doi:10.1007/s10614-020-10042-0.
7. Capuano N, Fenza G, Loia V, Stanzione C. Explainable artificial intelligence in cybersecurity: A survey. IEEE Access. 2022;10:93575-93600. doi:10.1109/ACCESS.2022.3204171.
8. Chen H, Gomez C, Huang CM, Unberath M. Explainable medical imaging AI needs human-centered design: Guidelines and evidence from a systematic review. npj Digit Med. 2022;5:156. doi:10.1038/s41746-022-00699-2.
9. Ghose P, Bhuiyan MRI, Hasan MN, Rakib SH, Mani L. Mediated and moderating variables between behavioral intentions and actual usages of fintech in the USA and Bangladesh through the extended UTAUT model. Int J Innov Res Sci Stud. 2025;8(2):113-125. doi:10.53894/ijirss.v8i2.5130.
10. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11):e745-e750. doi:10.1016/S2589-7500(21)00208-9.
11. Giudici P, Raffinetti E. Shapley-Lorenz eXplainable artificial intelligence. Expert Syst Appl. 2021;167:114104. doi:10.1016/j.eswa.2020.114104.
12. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. ACM Comput Surv. 2018;51(5):Article 93. doi:10.1145/3236009.
13. Hasan MN, Arman M, Bhuyain MMH, Chowdhury F, Bathula MK. Predictive analytics in healthcare: Strategies for cost reduction and improved outcomes in the USA. Int J Innov Res Sci Stud. 2025;8(8):142-150. doi:10.53894/ijirss.v8i8.10559.
14. Hasan MN, Bhuyain MMH, Chowdhury F. AI model that predicts antibiotic resistance from bacterial genomes using open sequencing data. Front Comput Sci Artif Intell. 2025;4(3). doi:10.32996/fcsai.2025.4.3.4.
15. Hasan MN, Bhuyain MMH, Chowdhury F, Arman M. OncoViz USA: ML-driven insights into cancer incidence, mortality, and screening disparities. J Med Health Stud. 2021;2(1):53-62. doi:10.32996/jmhs.2021.2.1.6.
16. Hasan MN, Miah MS, Ghose P, Jannat T, Bhuyain MMH, Chowdhury MSA, *et al*. A deep learning approach for brain tumor diagnosis: Combining an 8-layer CNN with rigorous K-fold validation. Math Model Eng Probl. 2025;12(12):4387-4396. doi:10.18280/mmep.121227.
17. Hasan MN, Papel MSI, Rasel IH, Akter S, Aktar MK, Abedin MZ, *et al*. Enhancing financial information security through advanced predictive analytics: A PRISMA-based systematic review. Edelweiss Appl Sci Technol. 2025;9(7):2222-2245. doi:10.55214/2576-8484.v9i7.9142.
18. Hasan MN, Rasel IH, Arman M, Ibrahim M, Jahan N. Strengthening U.S. financial and cybersecurity infrastructure with AI-driven fraud detection and risk analytics. J Comput Anal Appl. 2025;31(2):15-32.
19. Hasan MN, Rasel IH, Rahman M, Islam K, Arman M, Jahan N. Securing U.S. healthcare infrastructure with machine learning: Protecting patient data as a national security priority. Int J Comput Exp Sci Eng. 2022;8(3). doi:10.22399/ijcesen.3987.
20. Kost L, Lier SK, Breitner MH. An explainable artificial intelligence feature selection framework for transparent, trustworthy, and cost-efficient energy forecasting. Energy AI. 2025;22:100648. doi:10.1016/j.egyai.2025.100648.
21. Kuzlu M, Cali U, Sharma V, Güler Ö. Gaining insight into solar photovoltaic power generation forecasting utilizing explainable artificial intelligence tools. IEEE Access. 2020;8:187814-187823. doi:10.1109/ACCESS.2020.3031477.
22. Langer M, Oster D, Speith T, Hermanns H, Kästner L, Schmidt E, *et al*. What do we want from explainable artificial intelligence (XAI)? A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. Artif Intell. 2021;296:103473. doi:10.1016/j.artint.2021.103473.
23. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, *et al*. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2:56-67. doi:10.1038/s42256-019-0138-9.
24. Machlev R, Heistrene L, Perl M, Levy KY, Belikov J, Mannor S, *et al*. Explainable artificial intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. Energy AI. 2022;9:100169. doi:10.1016/j.egyai.2022.100169.
25. Miller T. Explanation in artificial intelligence: Insights from the social sciences. Artif Intell. 2019;267:1-38. doi:10.1016/j.artint.2018.07.007.
26. Nauta M, Trienes J, Pathak S, Nguyen E, Peters M, Schmitt Y, *et al*. From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable AI. ACM Comput Surv. 2023;55(13s):Article 295. doi:10.1145/3583558.
27. Neupane S, Ables J, Anderson W, Mittal S, Rahimi S, Banicescu I, *et al*. Explainable intrusion detection systems (X-IDS): A survey of current methods, challenges, and opportunities. IEEE Access. 2022;10:112392-112415. doi:10.1109/ACCESS.2022.3216617.
28. Noor Uddin Milon M, Ghose P, Chowdhury Pinky T, Nowshin Tabassum M, Nazmul Hasan M, Khatun M. An in-depth PRISMA-based review of cybercrime in a developing economy: Examining sector-wide impacts, legal frameworks, and emerging trends in the digital era. Edelweiss Appl Sci Technol. 2024;8(4):2072-2093. doi:10.55214/25768484.v8i4.1583.
29. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, *et al*. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.
30. Rasel IH, Arman M, Hasan MN, Bhuyain MMH. Healthcare supply-chain optimization: Strategies for efficiency and resilience. J Med Health Stud. 2022;3(4):171-182. doi:10.32996/jmhs.2022.3.4.26.
31. Rjoub G, Bentahar J, Abdel Wahab O, Mizouni R, Song A, Cohen RS, *et al*. A survey on explainable artificial intelligence for cybersecurity. IEEE Trans Netw Serv Manag. 2023;20(4):5115-5140. doi:10.1109/TNSM.2023.3282740.
32. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1:206-215. doi:10.1038/s42256-019-0048-x.
33. Samek W, Montavon G, Lapuschkin S, Anders CJ,

Müller KR. Explaining deep neural networks and beyond: A review of methods and applications. Proc IEEE. 2021;109(3):247-278. doi:10.1109/JPROC.2021.3060483.
34. Shadi MR, Mirshekali H, Shaker HR. Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects. Renew Sustain Energy Rev. 2025;216:115668. doi:10.1016/j.rser.2025.115668.
35. Zhang Z, Al Hamadi H, Damiani E, Yeun CY, Taher F. Explainable artificial intelligence applications in cyber security: State-of-the-art in research. IEEE Access. 2022;10:93104-93139. doi:10.1109/ACCESS.2022.3204051.
36. Zhou Y, Li H, Xiao Z, Qiu J. A user-centered explainable artificial intelligence approach for financial fraud detection. Finance Res Lett. 2023;58:104309. doi:10.1016/j.frl.2023.104309.

How to Cite This Article
Vasileva DA, Reynolds SM. Explainable Artificial Intelligence and Predictive Analytics: Applications Across Healthcare, Finance, Cybersecurity, and Sustainable Systems. Int J Med All Body Health Res. 2026;7(3):77-87. doi:10.54660/IJMBHR.2026.7.3.77-87.