British Journal of Multidisciplinary Studies
ISSN: 2978-8110
DOI: 10.32996/bjmss
Journal Homepage: www.al-kindipublisher.com/index.php/jbmss

BJMS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| REVIEW ARTICLE

Trustworthy and Explainable AI Across Critical Sectors: From Medical Diagnosis to Cyber-Physical and Business Systems

FNU Nurujjaman
College of Graduate and Professional Studies, Trine University, University Ave, Angola, IN 46703, USA
Corresponding Author: FNU Nurujjaman, **E-mail**: nadim142@gmail.com

| ABSTRACT

Artificial intelligence is increasingly deployed in critical decision environments, healthcare, assistive technologies, cyber-physical systems, agriculture, business analytics, cybersecurity, and distributed infrastructure, were inaccuracy, opacity, or unreliability may cause severe harm. While predictive accuracy has driven much of the field's progress, it is now broadly recognized as insufficient: trustworthy AI must also be explainable, robust, privacy-preserving, secure, fair, accountable, and practically deployable under real-world constraints. This structured critical review synthesizes using a six-axis taxonomy comprising critical sector, data modality, architecture family, explainability function, and trustworthiness concern. The review identifies six critical sectors, healthcare and biomedical AI, human-centered and assistive AI, cyber-physical systems and infrastructure, agriculture and sustainability, business and enterprise analytics, and cybersecurity and distributed intelligence—and eight architecture families, from conventional machine learning and CNNs through vision transformers, graph neural networks, Bayesian physics-guided models, generative AI, and federated learning systems. Synthesis reveals that while explainability mechanisms—visual, attentional, post-hoc, and knowledge-structured, are increasingly integrated across sectors, they are rarely validated against formal standards or shown to support trustworthy human oversight in deployment. Key gaps include explanation validation protocols, cross-sector benchmarking, privacy-preserving inference at scale, uncertainty quantification, and governance-aligned reporting. A structured research agenda is proposed that prioritizes validated explainability, federated deployment, robustness under distribution shift, fairness, and evidence maturity across all critical sectors.

| KEYWORDS

Trustworthy AI, Explainable AI, XAI, Vision transformers, Federated learning, Critical decision support, Governance, Cross-sector AI taxonomy.

| ARTICLE INFORMATION

ACCEPTED: 01 April 2026 **PUBLISHED:** 0245 May 2026 **DOI:** 10.32996/bjmss.2026.4.2.2x

1. Introduction

The maturation of AI from research curiosity to an operational instrument in consequential decision environments has created a new class of scientific and governance requirements [1]. In healthcare, AI systems increasingly support or recommend diagnoses that directly affect patient outcomes. In cyber-physical systems [2], AI monitors infrastructure whose failure carries safety and financial consequences on a scale. In business and enterprise settings, AI drives credit decisions, supply chain strategies, and workforce management policies that affect individual livelihoods. In each of these contexts, the predictive performance of an AI model, however impressive in controlled evaluation, is a necessary but wholly insufficient criterion for trustworthy deployment.

Trustworthy AI is a multidimensional concept that encompasses explainability, robustness, privacy, security, fairness, uncertainty quantification, human oversight, governance accountability, and evidence maturity. Explainability occupies a central role because it mediates the relationship between a model's internal computation and the human decision-maker who must act on its output [3]. However, explainability is also the most frequently oversimplified component of trustworthiness: attention maps, saliency overlays, and post-hoc feature importance scores are routinely presented as explanations without formal validation of their fidelity, completeness, or usefulness to the intended audience [4]. The risk of explanation theater, AI systems that appear interpretable without being genuinely accountable, is particularly acute in high-stakes sectors [5].

Across the critical sectors examined in this review healthcare and biomedical AI [6, 7, 8], human-centered and assistive AI [9, 4, 10], cyber-physical and infrastructure systems [6, 11, 12], agriculture and sustainability [13, 14, 15], business and enterprise analytics [6, 16, 17], and cybersecurity and distributed intelligence [18, 19, 20], a consistent pattern emerges: architectural sophistication has advanced significantly, but the deployment-critical properties of trustworthiness and validated explainability remain inconsistently addressed. This review constructs a cross-sector taxonomy and evidence map to diagnose these gaps and propose a structured research agenda (See Figure 1).

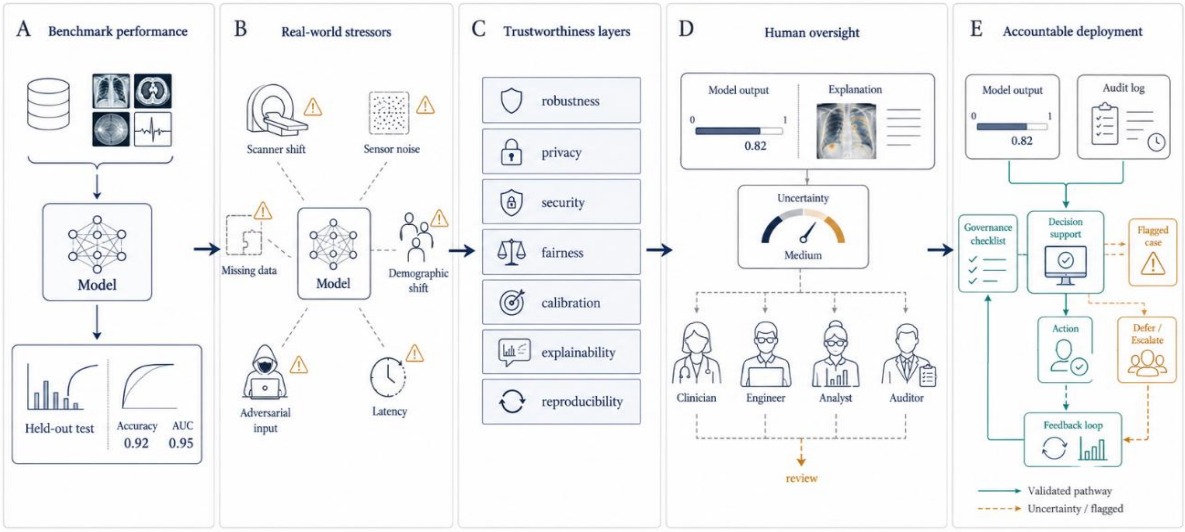

Figure 1. From accurate AI to trustworthy decision support.

2. Review Methodology

This review is presented as a structured critical review of a systematic review. Papers were assembled to provide balanced coverage of critical sectors, architecture families, data modalities, explainability functions, and trustworthiness concerns, enabling structured evidence mapping rather than exhaustive domain census.

A five-axis taxonomy organizes the corpus. Axis 1 classifies by critical sector: (i) healthcare and biomedical AI, (ii) human-centered, neuro-affective, and assistive AI, (iii) cyber-physical systems, IoT, robotics, and infrastructure, (iv) agriculture, environment, and sustainability, (v) business, enterprise, and organizational analytics, and (vi) cybersecurity, privacy, and distributed intelligence. Axis 2 classifies by data modality—medical images, facial and affective signals, EEG and physiological signals, IoT and sensor streams, acoustic-emission and industrial signals, text and natural language, graph and knowledge-structured data, business and tabular data, and multimodal data. Axis 3 identifies the architecture family: conventional machine learning, CNN-based deep learning and transfer learning, vision transformers and attention-based models, graph neural networks and knowledge graphs, hybrid and ensemble systems, Bayesian and physics-guided models, generative and agentic AI, and federated/edge/privacy-preserving systems. Axis 4 records the explainability function served: visual explanation, feature-level explanation, attention-based explanation, post-hoc interpretability, model transparency, knowledge-graph reasoning, human-readable decision support, and auditability. Axis 5 catalogues the trustworthiness concern: robustness, privacy, security, scalability, real-time feasibility, human oversight, governance, safety and accountability, fairness and access, and reproducibility and evidence maturity.

This taxonomy enables both vertical analysis—how architecture choices shape explainability and trust within a sector—and horizontal analysis—which trustworthiness concerns recur systematically across sectors. The full cross-sector evidence map is provided in Section 6.

3. Conceptual Foundations of Trustworthy and Explainable AI

3.1 From Accuracy-Centered AI to Trustworthy AI

The historical dominance of accuracy as the primary evaluation metric for AI systems reflects the origins of machine learning in benchmark competitions and controlled experiments. In critical deployment settings, however, accuracy on a held-out test set addresses only one of many requirements. A clinical decision support system for heart disease prediction [8] that achieves high accuracy on a training dataset but cannot generalize across patient demographics, handle missing values gracefully, or provide intelligible explanations to clinicians is not trustworthy, regardless of its benchmark performance. The transition from accuracy-centered to trustworthy AI requires integrating robustness to distribution shift, uncertainty quantification, privacy compliance, security against adversarial manipulation, fairness across protected groups, and governance-aligned reporting into the model development pipeline from the outset, not as post-hoc additions.

3.2 Explainability as a Decision-Support Requirement

Explainability is not a single property but a family of functions that serve different stakeholders. Explanations for model developers support debugging and architecture improvement. As shown in Figure 2, explanations for domain expert clinicians, agronomists, security analysts, must convey evidence in domain-relevant terms that support professional judgment. Explanations for end users must be accessible, concise, and actionable. Explanations for governance serve audit, compliance, and accountability functions. Post-hoc methods such as SHAP, LIME, and gradient-based saliency are the most widely used explanation approaches in the surveyed corpus, appearing in medical imaging [7, 21, 22], agricultural disease detection [13, 14, 15], and affective computing [10] applications. Attention mechanisms in transformer architectures offer a more architecturally integrated form of explanation [3, 23–25] but attention weights do not constitute complete causal explanations and should not be interpreted as such without validation. Knowledge-graph and NLP-based reasoning [16, 26] provides a structurally different form of explainability, traceable, entity-linked reasoning that is particularly suited to domains where relational context matters.

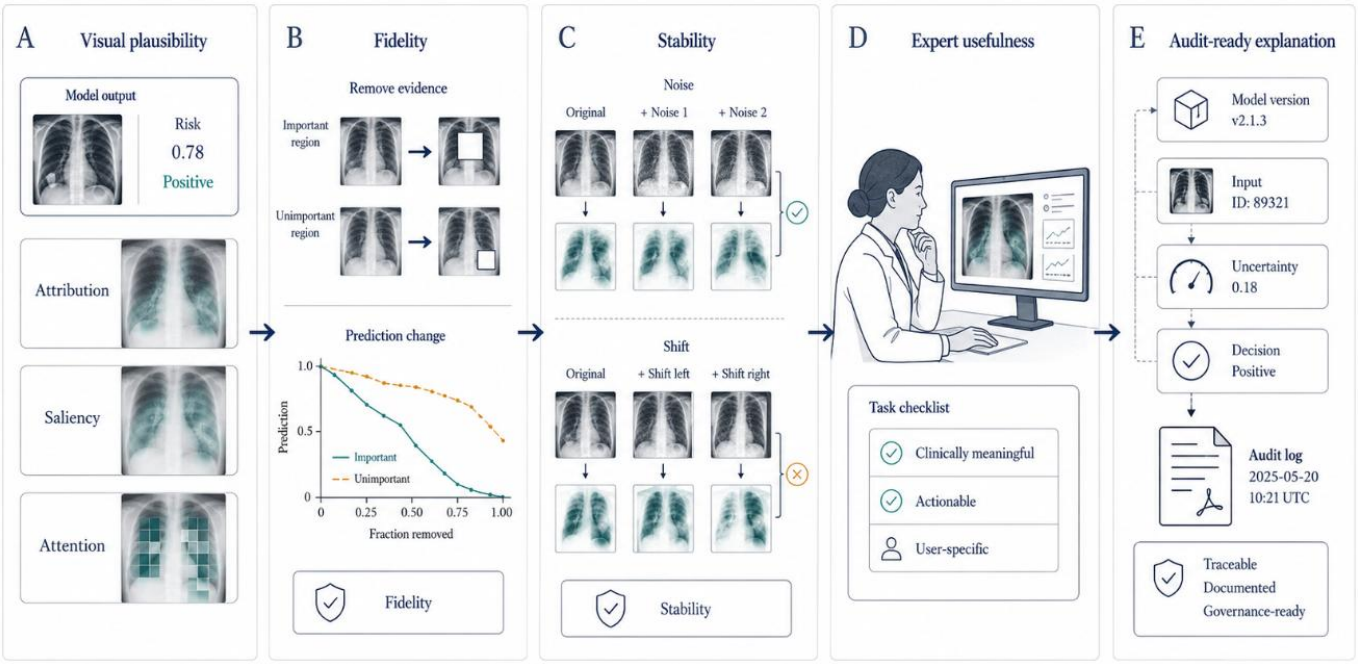

Figure 2. From visual explanation to validated explainability.

3.3 Trustworthiness Beyond Explainability

Explainability is a necessary but insufficient component of trustworthiness. As shown in Table 1, trustworthy AI framework for high-stakes decision support [5] provides a cross-sector conceptualization that positions explainability alongside robustness, privacy, security, fairness, and governance as co-equal requirements. Robustness concerns the preservation of model reliability under distribution shift, adversarial perturbation, and sensor degradation. Privacy concerns the protection of individual data during model training, inference, and deployment. Security addresses the vulnerability of AI systems to adversarial attacks, data poisoning, and model extraction. Fairness addresses differential treatment of individuals or groups, particularly in high-consequence settings such as credit scoring [27] and clinical screening [28]. Uncertainty quantification, addressed architecturally through Bayesian models [12]—enables AI systems to signal their own limitations, supporting human oversight by flagging predictions on which the system lacks confidence. Reproducibility and evidence maturity address the consistency of results across datasets, validation protocols, and deployment environments.

Table 1. Trustworthiness evaluation matrix across the AI lifecycle.

Lifecycle stage	Key objective	Minimum evidence	Main risk if omitted
Data readiness	Ensure representative and leakage-safe data.	Dataset source, sample size, class balance, labels, preprocessing, split strategy.	Inflated performance, bias, weak generalization.
Model development	Align model design with use context.	Architecture, training protocol, hyperparameters, compute, interpretability plan.	Overfitting, poor reproducibility, impractical complexity.

Lifecycle stage	Key objective	Minimum evidence	Main risk if omitted
Internal validation	Test reliability under controlled conditions.	Held-out performance, confusion matrix, calibration, confidence intervals.	Accuracy-only reporting and hidden class-level errors.
External validation	Assess robustness beyond the development data.	Independent dataset, subgroup results, distribution-shift and stress tests.	Fragile deployment and poor transferability.
Explainability validation	Verify that explanations are faithful and useful.	XAI method, target user, fidelity, stability, expert review.	Misleading explanations and false trust.
Privacy and security	Protect sensitive data and adversarial integrity.	Privacy mechanism, access control, threat model, security testing.	Data leakage, model compromise, regulatory risk.
Deployment feasibility	Confirm workflow and resource suitability.	Intended use, latency, hardware, usability, human-in-the-loop pathway.	Technically strong but unusable system.
Governance and monitoring	Support accountability after deployment.	Model versioning, audit trail, monitoring plan, update policy.	Silent performance drift and weak accountability.

3.4 Sector-Specific Trust Requirements

Trust requirements vary systematically across sectors. In healthcare, clinical interpretability, regulatory compliance, privacy, and accountability to patients and clinicians are primary. In cyber-physical and infrastructure systems, real-time reliability, safety certification, and fault explanation are paramount, a misclassified gas-pipeline fault [11, 29] or a miscalibrated wind-turbine sensor [12] may have immediate physical consequences. In business and enterprise settings, governance, audit trails, fairness, and strategic transparency are the dominant trust concerns [17, 30]. In agriculture, explanations must be actionable by field users with varying technical literacy, and models must be lightweight enough for edge deployment [14, 15]. In cybersecurity, security against adversarial manipulation, auditability, and real-time threat response are essential [15, 49, 74]. In assistive and human-centered AI, ethical oversight, personalization, accessibility, and sensitivity to vulnerable user populations add additional trust dimensions [4, 5, 9].

4. Architecture Families for Trustworthy and Explainable AI

4.1 Conventional Machine Learning and Structured Decision Models

Conventional machine learning, random forests, gradient boosted trees, logistic regression, and support vector machines—retains significant relevance in critical sectors, primarily because its feature-level explainability is well understood and its deployment footprint is manageable. In healthcare, structured patient data models [8] allow clinical decision support with feature importance outputs compatible with clinical reasoning. In business analytics, ML models for credit scoring [27], market trend forecasting [31], retail demand forecasting [32], project risk prediction [33], e-commerce pricing optimization [34], and small-business management [35] represent the operational backbone of enterprise AI decision support. Sentiment analysis of drug reviews [36] and Bengali social media [37] illustrate text-based ML in human-centered contexts. The explainability advantage of conventional ML is real but bounded: global feature importance does not guarantee local decision-level accountability, and structured models may embed historical biases that affect fairness in credit and hiring contexts [38].

4.2 CNN-Based Deep Learning and Transfer Learning

CNNs remain the dominant architecture for image-based critical AI, with transfer learning addressing the data-scarcity problem pervasive in medical, agricultural, and industrial datasets. Explainable AI-driven hybrid deep learning for skin cancer diagnosis [7] exemplifies the combination of CNN feature extraction with post-hoc explanation methods. Early leukemia diagnostics using image processing and transfer learning [39] and transfer learning for sleep stage classification under limited data [40] illustrate the domain adaptability of pre-trained CNN feature extractors. Facial emotion recognition systems, including a bidirectional Elman neural network [41] and a hybrid deep belief optimization system [42], extend convolutional feature learning to affective computing. Lightweight deep learning for concrete crack characterization via acoustic-emission signals [43] demonstrates that CNN architectures can be compressed for real-time edge deployment in industrial settings. Across all these applications, the

visual explanation methods used to interpret CNN decisions, gradient-weighted class activation mapping, occlusion sensitivity, provide useful indicators but do not constitute validated causal explanations, particularly for safety-critical applications.

4.3 Vision Transformers and Attention-Based Architectures

Vision transformers (ViTs) and their attention-based derivatives have emerged as a principal architecture family in high-stakes image classification, with the self-attention mechanism frequently cited as an explainability advantage over CNN-based approaches. The hybrid ViT for lung cancer diagnosis [3] and the hybrid ViT for prostate cancer in MRI [44] demonstrate the architecture's competitive performance in medical imaging. Hierarchical Swin Transformer ensembles for breast cancer [24] and Swin Transformer–driven cervical cell classification with web deployment [25] show that transformer architectures can be integrated with ensemble strategies and deployed via web interfaces. The dual-branch visual transformation model for ASD classification [45] and ASDnet [9] extend ViTs to facial and affective analysis. In precision agriculture, MaizeFormerX, a lightweight cross-scale attention ViT [14]—and the MaxViT model for soybean disease [46] demonstrate efficiency advances that begin to close the resource gap with CNNs. Global–local attention modeling for kidney disease classification from CT images [23] illustrates the diagnostic value of combining coarse global context with fine-grained local feature attention. It is critical to note that attention maps, while useful as visual communication tools, do not have the formal properties of causal explanation and should not be presented to clinicians, regulators, or auditors as complete justifications for model decisions without additional validation.

4.4 Hybrid, Ensemble, and Multimodal Fusion Systems

Hybrid and ensemble systems are motivated by both performance and explainability: combining heterogeneous learners improves generalization while enabling richer post-hoc explanation strategies. The explainable deep stacking ensemble for brain tumor diagnosis [22] and the stacking ensemble with XAI for cervical cancer [47] illustrate how ensemble diversity is leveraged alongside post-hoc explanation for clinical transparency. The ensemble transformer with post-hoc explanations for depression emotion and severity detection [10] extends this pattern to affective computing, where label ambiguity makes ensemble uncertainty particularly valuable. Multimodal fusion introduces additional explanation complexity: the hybrid multi-modal emotion recognition framework using InceptionV3DenseNet [58] and the vision-audio multimodal object recognition system using hybrid tensor fusion [48] must explain not only what a model predicted but which modality drove the prediction and how cross-modal interactions were resolved. The ViX-MangoEFormer ensemble for mango disease recognition with XAI [13] demonstrates that explainability can be maintained in stacking architectures through appropriate post-hoc design.

4.5 Graph Neural Networks and Knowledge-Graph Reasoning

Graph-structured representations offer a qualitatively different form of explainability from visual or feature-based methods: relational, traceable, and entity-linked. The GNN-enhanced gas-pipeline monitoring system [11] models propagate across sensor networks, enabling fault localization with structurally interpretable reasoning. Knowledge-graph and NLP integration for heuristic reasoning support [16] and the AddManBERT knowledge-graph construction for additive manufacturing design [26] demonstrate that BERT-based language models and knowledge graphs can be coupled to produce explanations that follow entity-relationship chains auditable by domain experts. This form of explanation structured and symbolic is particularly relevant in industrial, engineering, and enterprise settings where causal and relational accountability is required. The limitation is scalability: knowledge graphs require expert curation and may not generalize across rapidly evolving domains.

4.6 Bayesian, Physics-Guided, and Uncertainty-Aware Models

The capacity to express uncertainty is a defining property of trustworthy AI in safety-critical systems. The physics-guided Bayesian neural network for sensor fault detection in wind turbines [12] represents one of the most principled approaches in the corpus: by embedding physical system prior into the network, the model generates calibrated uncertainty estimates that can support human oversight decisions about maintenance and shutdown. Bayesian deep learning more broadly enables the distinction between epistemic uncertainty (arising from insufficient data) and aleatoric uncertainty (arising from inherent data noise), both of which are relevant to high-stakes decision support. The current corpus underrepresents this architecture family, reflecting a maturity gap between Bayesian methodology and domain-specific deployment. Addressing this gap is particularly urgent in medical imaging, infrastructure monitoring, and autonomous systems contexts.

4.7 Generative, Agentic, and Enterprise AI

Generative AI and agentic systems introduce new and qualitatively distinct trustworthiness challenges shown in Figure 3. Generative AI in enterprise information systems for transforming business intelligence [17] embeds large language model capabilities into strategic decision workflows, raising accountability questions that extend beyond standard ML governance frameworks: hallucination, factual unreliability, inconsistent chain-of-reasoning, and the absence of auditable computation logs.

Automated risk assessment and collaborative AI in agile project management [2] and AI for risk and decision in agile IT projects [49] exemplify agentic AI, in which systems initiate and coordinate decision workflows rather than merely responding to queries. AI-driven business analytics for IT strategy [50] and AI-enabled management information systems for governance and decision automation [30] address the organizational embedding of these systems, where adoption of trustworthy governance frameworks is at least as important as architectural performance. Crucially, generative AI systems should not be represented as explainable in the same terms as discriminative models: their reasoning processes are substantially less amenable to post-hoc attribution methods, and new explanation and audit frameworks are urgently needed.

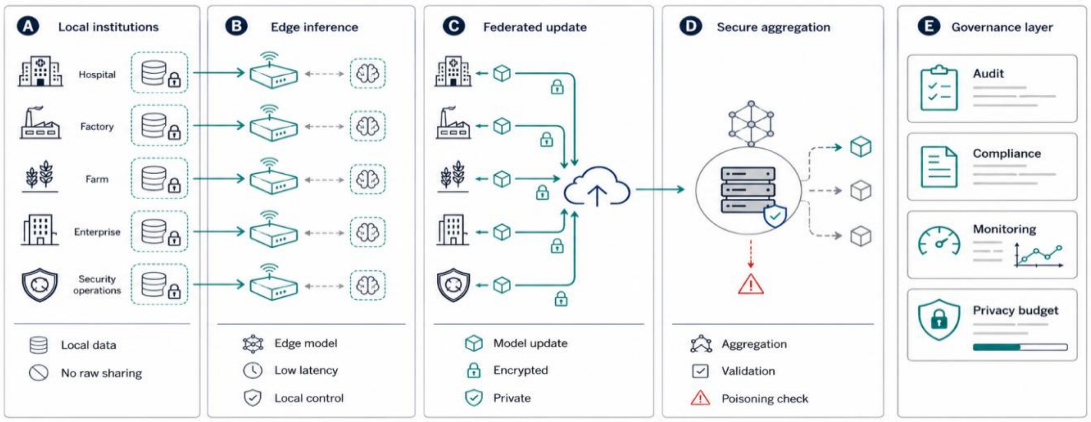

Figure 3. Federated and auditable AI deployment across critical sectors.

4.8 Edge-Cloud, Federated, Privacy-Preserving, and Distributed AI

The distributed intelligence and privacy-preserving deployment framework encompassing edge-cloud, 6G connectivity, and federated learning for secure and auditable decision support [19] represents the most comprehensive deployment architecture in the corpus. Federated learning addresses data sovereignty in healthcare and organizational analytics by enabling collaborative model training without centralizing raw data. Privacy-preserving behavior analytics for workforce retention [38] and the multimodal privacy-preserving cancer diagnosis framework [51] demonstrate operational privacy-preserving deployments. Stacking ensemble-based breast cancer classification with real-time web deployment [52] and Swin Transformer cervical cell screening with web interface [25] show that deployment-readiness—interface design, latency, cross-platform accessibility—must be treated as a first-class architectural constraint. The auditability requirement in [19] is critical: distributed systems that process sensitive data must generate auditable logs at each node, not just at the central aggregator, to maintain governance compliance.

5. Sector-Specific Synthesis

In healthcare, neural machine learning has been used to support stroke-risk assessment [82], while breast cancer diagnosis has been addressed through neural networks, dimensionality reduction, morphological feature analysis, and optimized neural architectures [80], [81]. Alongside performance improvement, explainable deep learning has been positioned as an important mechanism for making AI-based diagnostic decisions more transparent and clinically interpretable [85]. At the data-system level, privacy-first federated learning offers a distributed framework for scalable healthcare analytics without direct centralization of sensitive data [83]. Beyond clinical applications, AI-driven cybersecurity and digital twin technologies further demonstrate the relevance of intelligent systems in safeguarding essential infrastructure and enhancing predictive maintenance within industrial IoT environments [86], [84].

5.1 Healthcare and Biomedical AI

Healthcare constitutes the largest and most architecturally diverse sector in the corpus, reflecting both the intensity of AI research investment and the severity of deployment-trust requirements. Cancer diagnosis applications span skin cancer [7, 53], lung cancer [3, 79], breast cancer [24, 52], cervical cancer [25, 47], brain tumor [12], leukemia [39], prostate cancer [44], cytological cancer classification [21], and privacy-preserving multimodal cancer diagnosis [38]. The recurrence of explainability requirements across these applications, post-hoc XAI in [7], attention-based XAI in [3, 25, 24], comparative explainable ML in [21], and stacking-based transparency in [22, 47], reflects both clinical professional expectations and emerging regulatory requirements. Beyond oncology, kidney disease classification from CT images [23], Parkinson's screening via personalized voice-biomarker ML [28], sleep stage classification with transfer learning [40], heart disease prediction from structured data [8], diabetes management through AI-integrated healthcare information systems [54], depression severity detection [10], and the web-based deployment of screening tools [52] collectively illustrate the breadth of healthcare AI while underscoring the shared

deployment tension between clinical interpretability, privacy compliance, and generalization robustness. The multichannel analysis of imbalanced CT data for lung cancer [55] specifically highlights the class imbalance challenge that pervades medical imaging datasets and must be addressed for reliable clinical screening.

5.2 Human-Centered, Neuro-Affective, and Assistive AI

This sector addresses AI systems that support humans with communicative, cognitive, or affective needs—a context in which errors may directly harm vulnerable individuals and where explainability must be calibrated to user needs rather than technical audiences. ASD classification using facial grid-wise emotion features and dual-branch visual transformation [9, 45] and the ASD facial expression database [56] constitute a research cluster addressing a high-stakes developmental context. The AI-powered digital health platform for ASD students [4] extends this to therapeutic and educational personalization. Multimodal EEG analysis of neural synchrony using ML [57] and the standard tDCS model [58] address neuro-affective AI with direct clinical implications. Facial emotion recognition systems [41, 42, 59] and suicidal ideation detection using NLP and deep learning [60] occupy an ethically sensitive niche in which false negatives carry severe consequences and model confidence calibration is critical. The adaptive feedback system for learner improvement [61], the flex sensor hand glove for deaf and mute individuals [62], iris detection and recognition [63], and Bengali social media sentiment classification [37] extend the sector to accessibility, assistive communication, biometric security, and multilingual NLP. Online drug review sentiment extraction [36] bridges health and text-based human-centered AI. Across this sector, explainability for end users, not just model developers, and ethical oversight are the dominant trustworthiness requirements.

5.3 Cyber-Physical Systems, IoT, Robotics, and Infrastructure

Cyber-physical systems impose real-time reliability and safety requirements on AI that most evaluation frameworks do not address. IoT-based wireless battery monitoring for solar micro-grids [64], smart energy metering [65], and the smart healthcare medical box for elderly patients [66] represent AI-assisted monitoring in resource-critical and care-critical infrastructure. Wireless mesh network routing [67] and MANET routing protocol simulation [68] address network-layer decision support. HAPs communication systems optimization [1] extends infrastructure AI to airborne communication platforms, where real-time performance under varying channel conditions is operationally required. The question of full autonomy in underwater robotics [6] directly engages the human oversight axis: whether a system should operate without human supervision is not purely a technical question but one of safety accountability and governance. Gas-pipeline condition diagnosis via acoustic-emission imaging [29] and GNN-enhanced gas-pipeline monitoring [11] address safety-critical industrial infrastructure where fault localization must be accurate, explainable, and fast. Concrete crack characterization using lightweight DL and acoustic-emission signals [43], and sensor fault detection in wind turbines via physics-guided Bayesian neural networks [12], extend the sector to structural health monitoring and renewable energy infrastructure. Vision-audio multimodal object recognition [48] provides architecture-level evidence for multi-sensor fusion in infrastructure monitoring contexts.

5.4 Agriculture, Environment, and Sustainability

Agricultural AI decision support is characterized by lightweight deployment requirements, environmental variability, and the need for explanations accessible to field users. The maize leaf disease diagnosis system MaizeFormerX [14] and the MaxViT soybean disease model [46] represent the growing application of vision transformers to precision crop pathology, while the ViX-MangoEFormer ensemble for mango disease recognition with XAI [13] demonstrates that stacking ensemble transformers can maintain explainability under deployment constraints. The explainable transformer for cotton leaf diagnostics and fabric defect detection [15] and advanced deep learning for tea leaf disease precision [69] provide additional agricultural disease monitoring evidence. Lightweight ResNeXt for aquaculture disease diagnosis [70] illustrates the application of compressed CNN architectures to aquaculture monitoring, where inference speed and edge feasibility are operationally critical. AI-driven smart agriculture for crop yield and sustainability [71] and AI-driven solar financing for rural clinics and health small businesses [72] address the sustainability dimension, connecting agricultural AI to rural health infrastructure and financing resilience. The resilience-by-design AI framework [20] provides a cross-sectoral lens under which agricultural systems, health infrastructure, and security considerations are jointly addressed.

5.5 Business, Enterprise, and Organizational Decision Support

Business decision support AI spans the widest thematic range in the corpus and introduces the most heterogeneous trustworthiness requirements. Credit scoring for financially underserved businesses [27] raises fairness and access concerns: ML models trained on alternative data sources may perpetuate or amplify structural financial exclusion if not explicitly audited for disparate impact. Automated risk assessment in agile project management [2] and AI for IT project risk and decision [49] frame AI as a governance tool within organizational processes. Market basket analysis for healthcare service bundling [73] bridges health and business analytics. Blockchain and ML integration for supply chain management [74] introduces distributed ledger

trust mechanisms alongside predictive AI. Retail demand forecasting [32], small-business ML [35], e-commerce pricing [34], market trend forecasting [31], and customer satisfaction analytics in hospitality [75] constitute the operational forecasting and optimization cluster. The attention-enhanced deep learning system for business strategy optimization [76] applies transformer-based attention to enterprise decision support. Generative AI for enterprise business intelligence [17], digital transformation analytics [10], AI for IT strategy [50], agile IT risk AI [49], AI-ERP integration in dark factories [77], and AI-enabled management information systems for governance [30] address the strategic and governance layer of enterprise AI—the context in which trustworthiness has the broadest organizational consequences. Predictive analytics for project risk [33] and comprehensive small-business ML [35] complete the operational business analytics cluster.

5.6 Cybersecurity, Privacy, and Distributed Intelligence

Cybersecurity and privacy-preserving AI constitute both a standalone sector and a horizontal deployment requirement. The intelligent cybersecurity framework integrating ML-driven data protection and threat intelligence [18] addresses real-time threat detection and response in digital communications infrastructure. AI as a strategic engine for data security and digital resilience [78] and the resilience-by-design AI framework [20] position AI security within broader organizational and societal resilience architectures. Privacy-preserving behavior analytics for workforce retention [38] operationalizes differential privacy in organizational analytics. Trustworthy AI for high-stakes decision support across critical sectors [5] provides the overarching framework that defines trustworthiness in terms encompassing all six sectors reviewed here. The distributed intelligence and edge-cloud-6G-federated learning framework for secure and auditable decision support [19] represents the architectural frontier: integrating edge inference, cloud aggregation, 6G communication, and federated training into a system designed from the ground up for privacy, security, and auditability. As AI systems become more deeply embedded in digital communications and infrastructure, adversarial robustness, the ability to maintain reliable and trustworthy behavior under deliberate attack—becomes as important as robustness to natural distribution shift.

6. Trustworthiness and Explainability Challenges

6.1 Explanation Validity and Overinterpretation

A recurrent risk in the corpus is the use of attention maps and post-hoc saliency methods as proxies for complete model explanation. Attention weights in vision transformers [3, 14, 23, 24] indicate regions of the input that influenced the model's output but do not establish that these regions are causally decisive, clinically meaningful, or generalizable across inputs. Gradient-based saliency maps used to explain skin cancer [7] and skin lesion [53] classifiers may identify different regions depending on implementation details, making cross-system comparison unreliable. Post-hoc explanations for ensemble systems [10, 22, 47] face the additional challenge that the ensemble's collective decision may not be reducible to the explanation of any single base learner. The trustworthy AI framework [5] acknowledges these limitations and implies that explanation validity, the degree to which an explanation faithfully represents model reasoning must be evaluated independently from visual plausibility. Table 2 summarizes the minimum validation checks required before common XAI outputs can be treated as reliable explanations for high-stakes decision support.

Table 2. Explanation validation checklist by XAI method.

XAI method	What it explains	Required validation	Key limitation
Saliency / Grad-CAM	Image regions influencing prediction.	Fidelity test, perturbation test, stability under noise, expert review.	Visually plausible maps may not reflect true model reasoning.
Attention maps	Token, region, or feature attention patterns.	Attention-ablation test, consistency check, task relevance review.	Attention is not equivalent to causal explanation.
SHAP / LIME	Local feature contribution to individual predictions.	Feature-sensitivity test, repeated-run stability, subgroup comparison.	Results may vary with sampling, background data, or feature correlation.
Counterfactual explanation	Minimal input change needed to alter output.	Plausibility check, clinical/domain feasibility, robustness test.	Counterfactuals may be unrealistic or unsafe if domain constraints are ignored.
Feature importance	Global or local ranking of influential variables.	Cross-validation stability, subgroup audit, correlation assessment.	Important features may reflect bias or spurious associations.

XAI method	What it explains	Required validation	Key limitation
Knowledge graph reasoning	Entity relationships and rule-based inference paths.	Expert validation, relation accuracy check, audit-trail review.	Requires high-quality curated knowledge and may not scale easily.
Uncertainty-aware explanation	Confidence, ambiguity, and prediction limits.	Calibration curve, ECE, confidence interval, deferral analysis.	Poor calibration can create false reassurance.
Human-centered explanation	Whether users understand and act appropriately.	User study, task-completion test, decision-impact assessment.	Technically correct explanations may still be unusable.

6.2 Robustness and Distribution Shift

Distribution shifts the degradation of model performance when deployment conditions differ from training conditions is universally relevant but unevenly addressed. Medical imaging models face cross-scanner, cross-site, and cross-demographic shifts [44, 51, 55]. Agricultural models face seasonal and geographic shifts [46, 71, 79]. Infrastructure monitoring models must tolerate sensor degradation, environmental noise, and novel fault patterns [11, 12, 29]. Business forecasting models are vulnerable to economic regime changes [31, 32]. The physics-guided Bayesian neural network [12] addresses this directly by constraining model behavior through physical priority, offering robustness that purely data-driven models cannot guarantee. The resilience-by-design framework [20] provides a systemic perspective: robustness must be designed into systems at the architectural level, not retrofitted through post-training calibration.

6.3 Data Heterogeneity, Imbalance, and Uncertainty

High-stakes AI systems routinely encounter heterogeneous, imbalanced, and uncertain data. The multichannel analysis of imbalanced CT data for lung cancer [55] directly addresses the class imbalance problem in medical imaging, where rare but critical cases are underrepresented in training sets. Multimodal frameworks [38, 29, 58] introduce heterogeneity across modalities, requiring fusion strategies that do not allow one modality to dominate at the expense of another. EEG and physiological data [31] introduce temporal non-stationarity as a form of within-subject heterogeneity. Business and tabular data [4, 9] face missingness, categorical diversity, and distributional drift across time. Uncertainty quantification—architecturally addressed only in [57] within this corpus, is the principled response to data heterogeneity: models should express calibrated uncertainty rather than forced confidence when operating near decision boundaries.

6.4 Privacy, Security, and Federated Deployment

Privacy-preserving AI is both a regulatory requirement and an ethical obligation in health, workforce, and government contexts. The multimodal privacy-preserving cancer diagnosis framework [38] and privacy-preserving behavior analytics for workforce retention [45] demonstrate that utility and privacy can be simultaneously maintained, though with architecture-specific overhead. The distributed edge-cloud-6G federated learning framework [49] provides the most complete architectural response to privacy-preserving deployment, but introduces communication constraints, heterogeneous device capabilities, and potential for model poisoning attacks that must be addressed in the security layer. The intelligent cybersecurity framework [15] and AI-driven security analytics [74] address the threat intelligence layer, while the resilience-by-design framework [64] addresses systemic security under interdependent failure scenarios.

6.5 Real-Time Feasibility and Resource Constraints

Real-time inference under resource constraints is a deployment-critical requirement in IoT, agricultural, industrial, and clinical point-of-care contexts. Lightweight ResNeXt for aquaculture diagnosis [70], MaizeFormerX lightweight ViT [14], and lightweight DL for concrete crack characterization [43] explicitly optimize the accuracy-efficiency tradeoff for edge deployment. IoT-based solar micro-grid monitoring [64] and smart energy metering [65] embed AI inference in resource-constrained embedded hardware. Web-based deployment for cervical screening [25] and breast cancer diagnosis [52] demonstrates that cloud-hosted inference can satisfy real-time requirements, provided interface design and latency management are treated as first-class engineering concerns. HAPs communication optimization [1] and MANET routing simulation [68] address network-layer constraints that govern data transmission in distributed infrastructure systems.

6.6 Human Oversight, Accountability, and Governance

High-stakes AI systems should, by default, be designed to support human decision-making rather than replace a principle that has both ethical and legal dimensions. The question of full autonomy in underwater robotics [6] is explicitly framed as uncertain, reflecting the genuine difficulty of establishing the conditions under which unsupervised AI decision-making is responsible.

Automated risk assessment AI in agile project management [2] positions AI as a collaborator within governance-structured processes. The trustworthy AI framework [5] and AI-enabled MIS for governance [30] embed human oversight as a design requirement shown in Figure 5. Generative AI deployments [17] and agentic systems [2] introduce new accountability challenges: when an AI system initiates a decision workflow or generates strategic recommendations without a traceable reasoning chain, the locus of accountability becomes unclear. The adaptive feedback system for learners [61] and the ASD digital health platform [4] model responsible AI as systems that augment professional judgment rather than supplant it.

6.7 Benchmarking, Reproducibility, and Evidence Maturity

The corpus reveals significant heterogeneity in evaluation practices. Medical imaging studies report accuracy, sensitivity, and specificity on held-out test sets, but external validation on independent multi-site datasets is uncommon. Agricultural studies use domain-specific datasets infrequently shared across research groups, limiting replication. Business analytics studies rarely report confidence intervals, statistical significance, or uncertainty estimates. The comparative explainable ML analysis for cancer cytology [21] and the personalized Parkinson's screening model [28] represent relatively rigorous comparative evaluation designs, but neither can substitute for multi-institutional prospective validation. A governance-aligned evidence maturity framework analogous to clinical trial phases but adapted to AI decision-support systems is needed to provide researchers, practitioners, and regulators with a shared language for assessing deployment readiness across all critical sectors.

7. Future Research Directions

Several research priorities emerge from the cross-sector synthesis of trustworthy and explainable AI. First, the explanation validity gap remains a central limitation. Future studies should develop formal fidelity metrics and validation protocols for attention maps, saliency methods, and post-hoc XAI techniques [5,7]. These methods should be assessed not only by visual plausibility but also through explanation, fidelity scores, user-comprehension studies, and clinical or domain-specific usability trials. Second, there is a need for cross-sector trustworthy AI benchmarks. Current evaluation practices remain fragmented across healthcare, business, cyber-physical systems, agriculture, and cybersecurity. Future work should create shared multi-sector evaluation suits that jointly assess accuracy, robustness, fairness, privacy, and explainability. Such benchmarks could support multi-sector leaderboards, governance-compliance scoring, and deployment-readiness indices. Third, human-in-the-loop explainable AI should become a core design principle for high-stakes systems. AI models should include structured deferral mechanisms that activate when uncertainty is high or when explanations are inconsistent with domain knowledge [6,12]. Evaluation should measure decision quality with and without AI assistance, override frequency, user reliance, and downstream outcome tracking. Fourth, federated and privacy-preserving AI requires further development for multi-institutional and multi-sector deployment. Federated learning, edge-cloud inference, and privacy-preserving analytics should be scaled to settings where sensitive data cannot be centralized [24,38,51].

Key evaluation criteria should include privacy-budget consumption, model utility under federation, communication efficiency, and robustness against poisoning or adversarial updates. Fifth, stronger attention is needed for robustness and uncertainty quantification. Bayesian, physics-guided, and uncertainty-aware deep learning models should be integrated into safety-critical AI pipelines, particularly where deployment environments differ from training conditions [12,20]. These systems should be evaluated using calibration error, uncertainty coverage under distribution shift, out-of-distribution detection rate, and failure-case analysis. Sixth, lightweight and edge-deployable trustworthy AI should be prioritized for resource-constrained settings such as IoT, mobile health, agriculture, and industrial monitoring. Future studies should optimize vision transformers, ensemble models, and compact deep networks for embedded and real-time deployment [14,43,70]. Evaluation should report inference latency, memory footprint, energy demand, and the accuracy–efficiency Pareto frontier. Seventh, the field needs governance-aware reporting standards for AI decision-support systems. CONSORT- or TRIPOD-style reporting principles should be adapted to trustworthy AI studies across critical sectors [20,72]. Reporting should include dataset characteristics, validation design, subgroup analysis, calibration, explainability validation, privacy safeguards, deployment constraints, and governance documentation. Eighth, trustworthy generative and agentic AI requires dedicated audit frameworks. Unlike conventional predictive models, generative and agentic systems may produce recommendations, narratives, or workflow actions that are difficult to verify using standard attribution methods [2,17]. Future evaluation should measure factual accuracy, hallucination rate, audit-trail completeness, source traceability, and governance alignment. Ninth, evidence maturity frameworks should be established for high-stakes AI. Similar to clinical evidence hierarchies, AI systems should be classified from proof-of-concept models to externally validated, prospectively evaluated, and deployment-monitored systems. Evaluation should include evidence-level classification, external validation requirements, workflow testing, monitoring plans, and deployment-readiness indices. Finally, fairness, access, and accountability must be embedded into trustworthy AI evaluation. Future work should develop fairness-auditing protocols for systems that may have differential effects across demographic, geographic,

socioeconomic, or institutional groups [27,28,60]. Evaluation should include disparate-impact metrics, demographic parity gaps, subgroup calibration, fairness–accuracy trade-offs, and accountability mechanisms for affected users.

8. Limitations of the Review

The synthesis is thematic, architectural, explainability-oriented, and deployment-level in nature, rather than quantitative. It was not possible to extract specific performance metrics, dataset characteristics, sample sizes, validation protocols, or the specific XAI methods employed in each study. The review should be interpreted as a structured evidence map and taxonomic analysis rather than a quantitative meta-analysis. Full paper-level extraction, including access to methods, results, supplementary materials, and experimental details, would be required to support meta-analytic comparisons of explanation quality, dataset diversity, or validation rigor. Additionally, the curated corpus may not comprehensively represent all active threads in trustworthy AI research. Adjacent fields including legal AI, financial systemic risk, autonomous vehicles, and social welfare AI are not well represented and are acknowledged as important extensions of the present framework.

9. Conclusion

This structured critical review has synthesized across six critical sectors, healthcare and biomedical AI, human-centered and assistive AI, cyber-physical systems and infrastructure, agriculture and sustainability, business and enterprise analytics, and cybersecurity and distributed intelligence, using a five-axis taxonomy of sector, modality, architecture, explainability function, and trustworthiness concern. The synthesis reveals a landscape in which architectural sophistication is advancing rapidly, but the depth and validity of explainability and trustworthiness remain inconsistently addressed. Vision transformers, ensemble methods, graph neural networks, lightweight CNNs, and federated learning systems are each contributing to a new generation of capable decision-support tools. Yet the cross-sector view consistently reveals the same gaps: explanation methods are rarely validated for fidelity; robustness under distribution shift is rarely evaluated formally; uncertainty quantification is architecturally underrepresented; privacy-preserving deployment remains the exception; and governance-aligned reporting standards are absent from most evaluation frameworks.

The path forward requires treating trustworthy and explainable AI not as a compliance layer added to accurate models, but as a design requirement integrated from the earliest stage of system development. Validated explainability methods that can be audited by clinicians, regulators, field experts, and affected individuals; federated and privacy-preserving deployment pipelines that protect sensitive data across institutional boundaries; Bayesian and physics-guided uncertainty models that tell decision-makers when not to trust the AI; governance-aware reporting standards that document not only performance but evidence maturity, fairness, and deployment conditions—these are the foundations on which genuinely trustworthy AI across critical sectors must be built. Progress on these fronts will determine whether AI decision support fulfills its potential as a socially beneficial, accountable, and resilient technology for the decades ahead.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Adnan BM, Alamgir FM. Performance simulation and comparision in High Altitude Platforms (HAPs) communications systems under PSK, DPSK, QAM & FSK modulation schemes and AWGN, Rician & Rayleigh communication channels. 7th IEEE Annual Information Technology, Electronics and Mobile Communication Conference, IEEE IEMCON 2016. 2016 Nov 16. doi:10.1109/IEMCON.2016.7746080.

[2] Haque S, Chowdhury S, Faruq O, Akter R, Joy MSI, Munny MA, et al. Automated Risk Assessment and Collaborative Decision-Making AI Applications in Agile Project Management and Stakeholder Engagement. INTERNATIONAL JOURNAL OF ADVANCES IN SIGNAL AND IMAGE SCIENCES. 2026 Jan 15;12(1):915–23. doi:10.29284/V2JV8Q59.

[3] Debnath J, Uddin Khondakar Pranta AS, Hossain A, Sakib A, Rahman H, Haque R, et al. LMVT: A hybrid vision transformer with attention mechanisms for efficient and explainable lung cancer diagnosis. Inform Med Unlocked. 2025 Jan 1;57:101669. doi:10.1016/J.IMU.2025.101669.

[4] Haque S, Islam MS, Islam MI, Islam MS, Khan R, Tarafder MTR, Mohammad N. Enhancing adaptive learning, communication, and therapeutic accessibility through the integration of artificial intelligence and data-driven personalization in digital health platforms for students with autism spectrum disorder. Journal of Posthumanism. 2025;5(8):737-756. doi:10.63332/joph.v5i8.3255.

[5] Shakil MR, Hasan M, Tarek MIH, Polash FI, Meem EJ. Trustworthy AI for high-stakes decision support across critical sectors. World Journal of Advanced Engineering Technology and Sciences. 2026 Mar 31;18(3):241–53. doi:10.30574/WJAETS.2026.18.3.0152.

[6] Rohan A, Tolie HF, Hasan MJ, Kannan S. Full autonomy in underwater robotics systems: A realistic prospect? Eng Appl Artif Intell. 2025 Dec 22;162:112638. doi:10.1016/J.ENGAPPAI.2025.112638.

[7] Al Sakib A, Swapno SMR, Ahamed F, Mohiuddin A Bin, Bhuiyan MIH, Khan S, et al. Explainable AI-driven hybrid deep learning framework for accurate skin cancer diagnosis. Digit Health. 2026 Jan 1;12. doi:10.1177/20552076261438923.

[8] Rashid SU, Siddiqui MIH, Mahmud FU, Rahman MdS, Kabir AA, Shammah RS. Machine learning based clinical decision support for heart disease prediction using structured patient data. Journal of Computer Science and Technology Studies. 2024 Feb 25;6(1):340–50. doi:10.32996/JCSTS.2024.6.1.36.

[9] Alamgir FM, Zaman T, Hossain MS, Hassan MM, Alam MS. ASDnet: Classification model for individuals with autism spectrum disorder using facial grid-wise expressions features and dual-branch visual transformation. Biomed Signal Process Control. 2026 Jul 1;120:109999. doi:10.1016/J.BSPC.2026.109999.

[10] Islam S, Haque R, Khan MA, Mohiuddin AB, Siddiqui MIH, Limon ZH, Khushbu KG, Swapno SMMR, Ahmed MR, Appaji A. Ensemble transformer with post-hoc explanations for depression emotion and severity detection. iScience. 2026;29(2):114605. doi:10.1016/j.isci.2025.114605.

[11] Arifeen M, Hasan MJ, Rohan A, Kannan S, Prathuru A. Enhancing Acoustic Emission Driven Smart Gas-Pipeline Monitoring with Graph Neural Network. Springer Series in Advanced Manufacturing. 2025;Part F138:165–78. doi:10.1007/978-3-031-80154-9_8.

[12] Khan AAM, Rahman A, Mahmud FU, Bishnu KK, Nabil HR, Mridha MF, et al. A Physics-Guided Bayesian Neural Network for Sensor Fault Detection in Wind Turbines. IEEE Open Journal of the Computer Society. 2025;6:931–42. doi:10.1109/OJCS.2025.3577588.

[13] Noman A Al, Hossain A, Sakib A, Debnath J, Fardin H, Sakib A Al, et al. ViX-MangoEFormer: An Enhanced Vision Transformer–EfficientFormer and Stacking Ensemble Approach for Mango Leaf Disease Recognition with Explainable Artificial Intelligence. Computers 2025, Vol 14, Page 171. 2025 May 2;14(5):171. doi:10.3390/COMPUTERS14050171.

[14] Rahman MM, Gony MN, Ullah MS, Shuvra SMK, Haque R, Ahmed MdR, et al. MaizeFormerX: a lightweight vision transformer with cross-scale attention for explainable maize leaf disease diagnosis. Scientific Reports 2026 16:1. 2026 Mar 26;16(1):15160-. doi:10.1038/s41598-026-44550-0.

[15] Rahman Swapno SMM, Sakib A, Uddin Khondakar Pranta AS, Hossain A, Debnath J, Al Noman A, et al. Explainable Transformer Framework for Fast Cotton Leaf Diagnostics and Fabric Defect Detection. iScience. 2025 Feb 20;29(2):114411. doi:10.1016/j.isci.2025.114411.

[16] Haruna A, Noman K, Li Y, Makanda ILD, Zubair A, Hasand MJ, et al. Facilitating heuristic reasoning by utilizing knowledge graph and natural language processing. Knowl Based Syst. 2026 Feb 15;334:115153. doi:10.1016/J.KNOSYS.2025.115153.

[17] Haque S, Islam H, Sharmin F, Joy SI, Naher K, Rimi NN, et al. Generative Artificial Intelligence in Enterprise Information Systems: Transforming Business Intelligence and Strategic Decision Support Processes. International Journal of Interdisciplinary Cultural Studies, ISSN:2327-008XE-ISSN:2327-2554. 2025;20(2):1754–63. doi:10.18848/8P0S2E25.

[18] Shimu F, Faruq O, Azim KS, Chowdhury S, Shakirul M, Joy I, et al. INTELLIGENT CYBERSECURITY FRAMEWORK MACHINE LEARNING-DRIVEN DATA PROTECTION AND THREAT INTELLIGENCE INTEGRATION FOR MODERN DIGITAL COMMUNICATIONS. Int J Appl Math (Sofia). 2025 Oct 26;38(8s):620–32. doi:10.12732/IJAM.V38I8S.595.

[19] Shakil MR, Hasan M, Tarek MIH, Polash FI, Meem EJ. Distributed Intelligence and privacy preserving deployment: Edge–Cloud–6G–Federated Learning for Secure, Auditable Decision Support. World Journal of Advanced Engineering Technology and Sciences. 2026 Mar 31;18(3):268–79. doi:10.30574/WJAETS.2026.18.3.0154.

[20] Shakil MR, Hasan M, Imam M, Tarek H, Polash FI, Meem EJ. Resilience-by-Design: AI for security, sustainability and health in interdependent systems. World Journal of Advanced Engineering Technology and Sciences. 2026 Mar 31;18(3):254–67. doi:10.30574/WJAETS.2026.18.3.0153.

[21] Siddiqui MIH, Rahman MdS, Kabir AA, Mahmud FU, Rashid SU, Shammah RS. Comparative analysis of explainable machine learning models for cancer classification using cytological features. Journal of Medical and Health Studies. 2023 Oct 29;4(5):110–50. doi:10.32996/JMHS.2023.4.5.14.

[22] Haque R, Khan MA, Rahman H, Khan S, Siddiqui MIH, Limon ZH, et al. Explainable deep stacking ensemble model for accurate and transparent brain tumor diagnosis. Comput Biol Med. 2025 Jun 1;191:110166. doi:10.1016/J.COMPBIOMED.2025.110166 PubMed PMID: 40249992.

[23] Ahmed S, Miah MR, Shakil MR, Linkon AA, Siddiqui MIH, Malik AH. Global–Local Attention Modeling for Reliable Multiclass Kidney Disease Classification from CT Images. Journal of Medical and Health Studies. 2026 Mar 8;7(5):36–45. doi:10.32996/JMHS.2026.7.5.6.

[24] Ahmed MR, Rahman H, Limon ZH, Siddiqui MIH, Khan MA, Pranta ASUK, et al. Hierarchical Swin Transformer Ensemble with Explainable AI for Robust and Decentralized Breast Cancer Diagnosis. Bioengineering 2025, Vol 12, Page 651. 2025 Jun 13;12(6):651. doi:10.3390/BIOENGINEERING12060651.

[25] Shakil MR, Malik AH, Siddiqui MIH, Ahmed S, Miah MR, Linkon AA. Swin Transformer–Driven Cervical Cell Classification with Explainable AI and Web-Based Screening. Journal of Medical and Health Studies. 2026 Mar 8;7(5):25–35. doi:10.32996/JMHS.2026.7.5.5.

[26] Haruna A, Noman K, Li Y, Wang X, Hasan MJ, Alhassan AB. AddManBERT: A combinatorial triples extraction and classification task for establishing a knowledge graph to facilitate design for additive manufacturing. Advanced Engineering Informatics. 2025 Sep 1;67:103578. doi:10.1016/J.AEI.2025.103578.

[27] Mithun MM, Tanim SH, Tarannum R. Developing AI-Powered Credit Scoring Models Leveraging Alternative Data for Financially Underserved US Small Businesses. Repository Antis Publisher. 2025 Oct 18:699254.

[28] Ghosh BP, Shafiquzzaman Bhuiyan M, Kumar Bishnu K, Mahmud FU, Ray RK, Murshid M, et al. Personalized Machine Learning Models for Parkinson's Disease Screening via Voice Biomarkers: Accounting for Age, Gender, and Linguistic Variability. 2025.

[29] Hasan MJ, Noman K, Navid WU, Li Y, Haruna A, Ashfak K. Intelligent diagnosis of gas pipeline condition through multivariate analysis of acoustic emission signal-based imaging. Nondestructive Testing and Evaluation. 2025 Jan 31. doi:10.1080/10589759.2025.2456088.

[30] Shakil MR, Hasan M, Tarek MIH, Polash FI, Meem EJ. AI enabled management information systems for economic resilience and organizational performance: Analytics, governance, cyber risk and decision automation. World Journal of Advanced Engineering Technology and Sciences. 2026 Mar 31;18(3):294–307. doi:10.30574/WJAETS.2026.18.3.0156.

[31] Hossain MS, Khan A, Das P, Haque MSU, Kamruzzaman F, Akter S, et al. Enhanced market trend forecasting using machine learning models: a study with external factor integration. International Interdisciplinary Business Economics Advancement Journal. 2025 Jan 7;6(01):5–12. doi:10.55640/BUSINESS/VOLUME06ISSUE01-02.

[32] Shak S, Mozumder SA, Hasan A, Das AC, Miah R, Akter S. OPTIMIZING RETAIL DEMAND FORECASTING: A PERFORMANCE EVALUATION OF MACHINE LEARNING MODELS INCLUDING LSTM AND GRADIENT BOOSTING [Internet]. 2024. doi:10.37547/tajet/Volume06Issue09-09.

[33] Tanim SH, Ahmad S, Mithun MU, Tarannum R, Refat F, Sunny NM. Leveraging Predictive Analytics for Risk Identification and Mitigation in Project Management. Journal of Information Systems Engineering and Management. 2025 May 7;10(43s):1041–52. doi:10.52783/JISEM.V10I43S.8523.

[34] Chowdhury MS, Shak MS, Devi S, Miah MR, Mamun A Al, Ahmed E, et al. Optimizing E-Commerce Pricing Strategies: A Comparative Analysis of Machine Learning Models for Predicting Customer Satisfaction. The American Journal of Engineering and Technology. 2024 Sep 4;06(09):6–17. doi:10.37547/TAJET/VOLUME06ISSUE09-02.

[35] Naznin R, Sarkar MAI, Asaduzzaman M, Akter S, Mou SN, Miah MR, et al. ENHANCING SMALL BUSINESS MANAGEMENT THROUGH MACHINE LEARNING: A COMPARATIVE STUDY OF PREDICTIVE MODELS FOR CUSTOMER RETENTION, FINANCIAL FORECASTING, AND INVENTORY OPTIMIZATION. International Interdisciplinary Business Economics Advancement Journal. 2024 Nov 22;5(11):21–32. doi:10.55640/BUSINESS/VOLUME05ISSUE11-03.

[36] Haque R, Laskar SH, Khushbu KG, Hasan MJ, Uddin J. Data-Driven Solution to Identify Sentiments from Online Drug Reviews. Computers 2023, Vol 12, Page 87. 2023 Apr 21;12(4):87. doi:10.3390/COMPUTERS12040087.

[37] Haque R, Islam N, Tasneem M, Das AK. Multi-class sentiment classification on Bengali social media comments using machine learning. International Journal of Cognitive Computing in Engineering. 2023 Jun 1;4:21–35. doi:10.1016/J.IJCCE.2023.01.001.

[38] Tanim SH, Tarannum R, Mithun MMU. Privacy-preserving behavior analytics for workforce retention approach. American Journal of Engineering, Mechanics and Architecture. 2023;1(9):188-215.

[39] Haque R, Al Sakib A, Hossain MF, Islam F, Ibne Aziz F, Ahmed MR, et al. Advancing Early Leukemia Diagnostics: A Comprehensive Study Incorporating Image Processing and Transfer Learning. BioMedInformatics 2024, Vol 4, Pages 966-991. 2024 Apr 1;4(2):966–91. doi:10.3390/BIOMEDINFORMATICS4020054.

[40] Uddin Mahmud F, Rahman H, Hossain Limon Z, Alam Khan M, Bin Jashim F. Transfer learning approach for sleep stage classification with limited training data. International Journal of Science and Research Archive. 2025;2025(02):1469–79. doi:10.30574/ijsra.2025.15.2.1506.

[41] Alamgir FM, Alam MS. A Novel Deep Learning-Based Bidirectional Elman Neural Network for Facial Emotion Recognition. https://doi.org/101142/S0218001422520164. 2022 Aug 3;36(10). doi:10.1142/S0218001422520164.

[42] Alamgir FM, Alam MS. An artificial intelligence driven facial emotion recognition system using hybrid deep belief rain optimization. Multimedia Tools and Applications 2022 82:2. 2022 Jun 27;82(2):2437–64. doi:10.1007/S11042-022-13378-X.

[43] Habib MA, Hasan MJ, Kim JM. A Lightweight Deep Learning-Based Approach for Concrete Crack Characterization Using Acoustic Emission Signals. IEEE Access. 2021;9:104029–50. doi:10.1109/ACCESS.2021.3099124.

[44] Debnath J, Bin Mohiuddin A, Pranta ASUK, Sakib A, Hossain A, Shanto MM, et al. Hybrid Vision Transformer Model for Accurate Prostate Cancer Classification in MRI Images. 2025 International Conference on Electrical, Computer and Communication Engineering, ECCE 2025. 2025. doi:10.1109/ECCE64574.2025.11013952.

[45] Alamgir FM, Zaman T, Hassan MM, Jonayed MR, Alam MS. Classification Model for Autism Spectrum Disorder Individuals: Utilizing Facial Grid-Wise Emotion Features and Dual-Branch Visual Transformation. PEEIACON 2024 - International Conference on Power, Electrical, Electronics and Industrial Applications. 2024;864–9. doi:10.1109/PEEIACON63629.2024.10800506.

[46] Pranta ASUK, Fardin H, Debnath J, Hossain A, Sakib AH, Ahmed MR, et al. A Novel MaxViT Model for Accelerated and Precise Soybean Leaf and Seed Disease Identification. Computers 2025, Vol 14, Page 197. 2025 May 18;14(5):197. doi:10.3390/COMPUTERS14050197.

[47] Siddiqui MIH, Khan S, Limon ZH, Rahman H, Khan MA, Al Sakib A, et al. Accelerated and accurate cervical cancer diagnosis using a novel stacking ensemble method with explainable AI. Inform Med Unlocked. 2025 Jan 1;56:101657. doi:10.1016/J.IMU.2025.101657.

[48] Ahmed MR, Haque R, Rahman SMA, Reza AW, Siddique N, Wang H. Vision-audio multimodal object recognition using hybrid and tensor fusion techniques. Information Fusion. 2026 Feb 1;126:103667. doi:10.1016/J.INFFUS.2025.103667.

[49] Karshiboev A, Al-Samad K, Tarafdar R, Rimi NN, Islam S, Papel SI, et al. Artificial Intelligence For Risk And Decision Assessment In Agile IT Projects: A Thematic Analysis And Dynamic Structuration Framework Approach. INTERNATIONAL JOURNAL OF ADVANCES IN SIGNAL AND IMAGE SCIENCES. 2026 Feb 10;12(2s):387–410. doi:10.29284/9K2NX425.

[50] Haque S, Mohammad N, Mambetaliev A, Karshiboev A, Lucky KY, Khan MTH, Islam H. Artificial intelligence-driven business analytics for IT strategy: Advancing decision-making, real-time insights, and organizational agility through intelligent automation and data integration. Journal of Posthumanism. 2025;5(6):1848-1863. doi:10.63332/joph.v5i6.2287.

[51] Kabir AA, Mahmud FU, Rahman MdS, Rashid SU, Siddiqui MIH, Shammah RS. Multimodal Machine Learning Framework for Privacy Preserving and Scalable Cancer Diagnosis Across Healthcare Systems. Journal of Adaptive Learning Technologies. 2024 Jun 15;1(6).

[52] Jashim F Bin, Refat FR, Karim MH, Mahmud FU, Sakib AH. Stacking ensemble-based breast cancer classification: Enhancing diagnostic accuracy with deep learning and real-time web deployment. International Journal of Science and Research Archive. 2025 May 30;15(2):1417–31. doi:10.30574/IJSRA.2025.15.2.1502.

[53] Linkon AA, Shakil MR, Ahmed S, Miah MR, Malik AH. Explainable Transformer-Based Skin Lesion Classification from Clinical Images. Journal of Medical and Health Studies. 2026 Mar 8;7(5):46–55. doi:10.32996/JMHS.2026.7.5.7.

[54] Lucky KY, Haque S, Al-Samad K, Akter R, Faruq O, Azim KS, et al. AI-Powered Healthcare Information Systems Securing Diabetes Management Through Integrated Technology Solutions and Enhanced Patient Care Delivery. Vascular and Endovascular Review. 2025;8(11s):465-476.

[55] Sohaib M, Hasan MJ, Zheng Z. A multichannel analysis of imbalanced computed tomography data for lung cancer classification. Meas Sci Technol. 2024 May 7;35(8):085401. doi:10.1088/1361-6501/AD437F

[56] Alamgir FM, Saif SMH, Hossain SM, Al Hadi A, Alam MS. Facial Expression Database of Autism Spectrum Disorder Children. European Chemical Bulletin. 2023;12(Special Issue 4):21109-21120. doi:10.48047/ecb/2023.12.Si4.1851.

[57] Majumdar J, Apu MH, Rahman M, Zaman T, Alamgir FM, Hassan MdM. Multimodal EEG Analysis of Neural Synchrony in Minimal Phrase Processing Using Machine Learning. 2025 IEEE 4th International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON). 2025 Nov 27;296–301. doi:10.1109/RAAICON69033.2025.11502190.

[58] Sourav MSU, Rahman A, Al Mamun A, Alamgir FM. Standard transcranial direct current stimulation (tDCS) model. International Journal of Computer Networks and Communications Security. 2017;5(12):264-270.
[59] Alamgir FM, Alam MS. Hybrid multi-modal emotion recognition framework based on InceptionV3DenseNet. Multimedia Tools and Applications 2023 82:26. 2023 Mar 27;82(26):40375–402. doi:10.1007/S11042-023-15066-W.
[60] Haque R, Islam N, Islam M, Ahsan MM. A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning. Technologies 2022, Vol 10, Page 57. 2022 Apr 29;10(3):57. doi:10.3390/TECHNOLOGIES10030057.
[61] Qadir HM, Khan RA, Rasool M, Sohaib M, Shah MA, Hasan MJ. An adaptive feedback system for the improvement of learners. Scientific Reports. 2025;15:17242. doi:10.1038/s41598-025-01429-w.
[62] Al Mamun A, Alamgir FM. Flex Sensor Based Hand Glove for Deaf and Mute People. International Journal of Computer Networks and Communications Security [Internet]. 2017 [cited 2026 May 18];5(2):38–48. Available from: www.ijcncs.org.
[63] Biswas R, Uddin J, Hasan MJ. A New Approach of Iris Detection and Recognition. International Journal of Electrical and Computer Engineering (IJECE). 2017 Oct 1;7(5):2530–6. doi:10.11591/ijece.v7i5.pp2530-2536.
[64] Mahamud S, Hossain MS, Hassan MM, Maruf MY, Rafi MAH, et al. IoT based wireless battery monitoring system for enhanced solar micro-grid performance in Bangladesh. In: Arefin MS, Kaiser MS, Bhuiyan T, Based MA, Ray K, editors. Proceedings of the 3rd International Conference on Big Data, IoT and Machine Learning. BIM 2025. Lecture Notes in Networks and Systems, vol. 1798. Cham: Springer; 2026. p. 474-489. doi:10.1007/978-3-032-15346-3_33.
[65] Haque MM, Choudhury ZH, Alamgir FM. IoT Based Smart Energy Metering System for Power Consumers. ICIET 2019 - 2nd International Conference on Innovation in Engineering and Technology. 2019 Dec 23. doi:10.1109/ICIET48527.2019.9290661.
[66] Al-Mahmud O, Khan K, Roy R, Mashuque Alamgir F. Internet of Things (IoT) based smart health care medical box for elderly people. 2020 International Conference for Emerging Technology, INCET 2020. 2020 Jun 1. doi:10.1109/INCET49848.2020.9153994.
[67] Alamgir FM, Ahmed F, Miah M, Mohammad H, Barua S. A Novel Routing Algorithm for Inter-Group Load Balancing in Wireless Mesh Networks. 21st Saudi Computer Society National Computer Conference, NCC 2018. 2018 Dec 27. doi:10.1109/NCG.2018.8593192.
[68] Ahmed F, Alamgir FM. Simulation-Based Proportional Study of Routing Protocols for MANET. International Journal of Computer Networks and Communications Security [Internet]. 2017 [cited 2026 May 18];5(12):271–6. Available from: www.ijcncs.org.
[69] Zakirhossain M, Khan MM, Thapa S, Uddin R, Meem EJ, Niloy SK, et al. Advanced Deep Learning Techniques for Precision Diagnosis of Tea Leaf Diseases. 2025 IEEE International Conference on Emerging Technologies and Applications, MPSec ICETA 2025. 2025. doi:10.1109/MPSECICETA64837.2025.11118779.
[70] Masum AKM, Khan MFI, Mahmud FU, Hassan MM, Khaliluzzaman M. Improving aquaculture disease diagnosis with lightweight ResNeXt architectures. In: 2025 3rd International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings); 2025. doi:10.1109/AIBThings66987.2025.11296219.
[71] Riipa MB, Saha S, Ferdousmou J, Khatoon R, Mohammad N, Hossain M, et al. AI-Driven Smart Agriculture: Optimizing Crop Yield and Sustainability in the U.S. International Conference on Electrical, Computer, and Energy Technologies, ICECET 2025. 2025. doi:10.1109/ICECET63943.2025.11472088.
[72] Sakhawat Hussain T, Md Manarat Uddin M, Rahanuma T. Sustaining Vital Care in Disasters: AI-Driven Solar Financing for Rural Clinics and Health Small Businesses [Internet]. 2025 Sep 30 [cited 2026 May 18]. Available from: https://semantjournals.org/index.php/AJTA/article/view/2528.
[73] Rimon RH, Nurujjaman, Mithun MMU. Market basket analysis for healthcare services to identify bundled care offerings. Frontiers in Computer Science and Artificial Intelligence. 2025 Apr 25;4(3):44–67. doi:10.32996/FCSAI.2025.4.3.5.
[74] Rahman T, Uddin MK, Hosen MM, Bhattacharjee B, Taluckder MS, Mou SN, et al. BLOCKCHAIN APPLICATIONS IN BUSINESS OPERATIONS AND SUPPLY CHAIN MANAGEMENT BY MACHINE LEARNING. International Journal of Computer Science & Information System. 2024 Nov 15;9(11):17–30. doi:10.55640/IJCSIS/VOLUME09ISSUE11-03.
[75] Talukder T, Masud S Bin, Miah MdR, Hera A, Faruque MdO, Talukder T, et al. An Examination of How Social Media Participation and Customer Satisfaction Affect the Likelihood that a Business Will Make Another Transaction in the Hospitality Sector. Open Access Library Journal. 2025 Jan 27;12(1):1–15. doi:10.4236/OALIB.1112802.
[76] Mahmud FU, Rahman A, Khan MA, Bishnu KK, Eva AA, Maua J. FuseAttenX: Leveraging Attention-Enhanced Deep Learning for Business Strategy Optimization. 2025 IEEE 4th International Conference on Computing and Machine Intelligence, ICMI 2025 - Proceedings. 2025. doi:10.1109/ICMI65310.2025.11141140.
[77] Islam MS, Islam MI, Mozumder AQ, Khan MTH, Das N, Mohammad N. A Conceptual Framework for Sustainable AI-ERP Integration in Dark Factories: Synthesising TOE, TAM, and IS Success Models for Autonomous Industrial Environments. Sustainability (2071-1050). 2025 Oct 15;17(20):9234. doi:10.3390/SU17209234.
[78] Faruq O, Chowdhury S, et al. Artificial intelligence as the strategic engine of data security, analytics, and digital communication for a resilient digital future. Journal of Information and Knowledge Management. 2025;20(2):1764-1773.
[79] Rahman Swapno SMM, Sakib A, Uddin Khondakar Pranta AS, Hossain A, Debnath J, Al Noman A, et al. Explainable Transformer Framework for Fast Cotton Leaf Diagnostics and Fabric Defect Detection. iScience. 2025 Feb 20;29(2):114411. doi:10.1016/j.isci.2025.114411.
[80] Khan MA, Parveen R, Ahmed I, Milon MH, Khan TA. High-Accuracy Breast Cancer Diagnosis Using Neural Networks and Dimensionality Reduction Techniques. In2025 IEEE 19th International Conference on Open Source Systems and Technologies (ICOSST) 2025 Dec 1 (pp. 1-6). doi:10.1109/ICOSST69113.2025.11315291.
[81] Raja MR, Milon MH, Ahmed I, Papel MS, Khan MA, Islam MZ. Optimizing Neural Architectures for Accurate Diagnosis of Breast Cancer from Morphological Features. In2025 3rd International Conference on Cyber Resilience (ICCR) 2025 Jul 3 (pp. 1-6). doi:10.1109/ICCR67387.2025.11292567.
[82] Khan MA, Papel MS, Milon MH, Ahmed I, Islam MZ, Raja MR. Optimizing Stroke Prediction in Healthcare with Neural Machine Learning Algorithms. In2025 3rd International Conference on Cyber Resilience (ICCR) 2025 Jul 3 (pp. 1-7). doi:10.1109/ICCR67387.2025.11292555.
[83] Ahmed I, Papel MS, Raja MR, Islam MZ, Khan MA, Milon MH. Privacy-First Federated Learning Models for Scalable Healthcare Data Processing. In2025 3rd International Conference on Cyber Resilience (ICCR) 2025 Jul 3 (pp. 1-6). doi:10.1109/ICCR67387.2025.11291736.

[84] Milon MH, Rahman MM, Papel MS, Raja MR, Semi MM, Tarafder MT. Digital Twin Technology for Predictive Maintenance in Industrial Iot Environments: Enhancing Operational Efficiency and Asset Longevity. In2025 3rd International Conference on Business Analytics for Technology and Security (ICBATS) 2025 May 1 (pp. 1-8). doi:10.1109/ICBATS66542.2025.11258212.

[85] Siam MA, Ahmed I, Khan MA, Islam MA, Milon MH, Ahamed A, Islam MZ. Explainable Deep Learning Models for Medical Diagnosis: Bridging the Gap between AI and Healthcare. In2025 3rd International Conference on Business Analytics for Technology and Security (ICBATS) 2025 May 1 (pp. 1-7). doi:10.1109/ICBATS66542.2025.11258368.

[86] Islam MZ, Siam MA, Ahmed I, Khan MA, Islam MA, Milon MH. Fortifying Healthcare and Essential Infrastructure with AI-Driven Cybersecurity Technologies. In2025 International Conference on Metaverse and Current Trends in Computing (ICMCTC) 2025 Apr 10 (pp. 1-9). doi:10.1109/ICMCTC62214.2025.11196395.