Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTC
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| REVIEW ARTICLE

Multimodal and Hybrid Artificial Intelligence for Real-World Decision-Making: Methods, Evidence, and Applications

Md Risalat Hossain Ontor[1], Md Golam Sarwar[2], Shahadat Hossain[2]
[1] *Doctor of Management, International American University, 3440 Wilshire Blvd., 10th Floor, #1000, Los Angeles, CA 90010, USA*
[2] *Department of Information Systems, Pacific States University, 3530 Wilshire Boulevard, Suite 1110, Los Angeles, CA 90010, USA*
Corresponding Author: Shahadat Hossain, **E-mail**: shahadatremon8@gmail.com

| ABSTRACT
Real-world decision-making rarely depends on a single data stream. Healthcare diagnosis, industrial fault detection, agricultural disease monitoring, business intelligence, cybersecurity threat response, and assistive technology all require AI systems capable of integrating heterogeneous evidence from images, text, physiological signals, sensors, graphs, tabular records, and their combinations. Multimodal and hybrid AI systems address this challenge by combining complementary data modalities, complementary architectures, fusion strategies, domain knowledge, and deployment infrastructures. This review identifies and critically examines four categories of directly multimodal or hybrid evidence, multimodal EEG analysis, vision-audio fusion, hybrid multimodal emotion recognition, and privacy-preserving multimodal cancer diagnosis, alongside a broader set of hybrid, ensemble, attention-based, graph-guided, and Bayesian architectures that advance multimodal integration. It situates these within seven application domains and examines the cross-cutting challenges of fusion design, modality alignment, interpretability, robustness, privacy, computational feasibility, and human oversight. Synthesis reveals that while fusion strategies have diversified from feature concatenation through tensor and attention-based fusion to knowledge-guided integration, evidence validation practices, including fusion ablation, modality-dropout testing, and calibrated uncertainty reporting, remain inconsistently applied. A structured research agenda addresses these gaps with eleven actionable future directions.

| KEYWORDS

Multimodal AI, Hybrid AI, Fusion strategies, Explainable AI, Decision support systems, Trustworthy AI, Federated learning, Cross-domain AI taxonomy

| ARTICLE INFORMATION

ACCEPTED: 09 April 2026 **PUBLISHED:** 22 May 2026 **DOI:** 10.32996/jcsts.2026.8.7.10

1. Introduction

Consequential decisions in healthcare, industry, business, and public infrastructure are rarely made on the basis of a single data type. A clinician diagnosing cancer considers imaging evidence, pathological reports, patient history, and laboratory values simultaneously. An industrial maintenance engineer interprets vibration spectra, acoustic-emission patterns, thermal images, and equipment logs together. A business analyst integrates transactional records, social media signals, market indicators, and operational data. The heterogeneity of real-world evidence motivates the development of multimodal and hybrid AI systems: architectures that can process, align, and integrate information across different data modalities, representation spaces, and knowledge structures to produce more reliable and complete decision support than any single-modality system could provide. The literature on multimodal AI spans a wide range of fusion strategies, early data-level concatenation, intermediate feature-level fusion, attention-based cross-modal alignment, tensor product fusion, ensemble and stacking decision fusion, and knowledge-guided symbolic integration, applied across medical imaging, physiological signal analysis, natural language processing, sensor networks, graph-structured data, and business analytics. Hybrid architectures, combinations of CNNs, transformers, graph neural networks, conventional ML, Bayesian models, and ensemble systems, extend this diversity further by combining the representational strengths of multiple model families within a single decision pipeline.

This review examines multimodal and hybrid AI from three complementary perspectives. First, it characterizes the methodological foundations of fusion design, hybrid architecture construction, and their deployment implications. Second, it maps the evidence across seven application domains using a seven-axis taxonomy. Third, it critically evaluates the challenges modality alignment, fusion validity, interpretability, robustness, privacy, computational feasibility, and governance that must be resolved for multimodal and hybrid AI to move from research demonstrations to trustworthy real-world deployment. Papers representing direct multimodal evidence [1, 14, 48, 71], hybrid architecture evidence [11, 20, 27, 37, 43, 54, 65, 74, 78, 79], and the broader domain and modality context [4, 13, 38, 62] are synthesized within this framework. Figure 1 shows the conceptual pathway through which heterogeneous evidence sources are transformed into modality-specific representations, integrated through fusion or reasoning mechanisms, and translated into accountable decision support with uncertainty estimation, explanation, human review, and post-deployment monitoring.

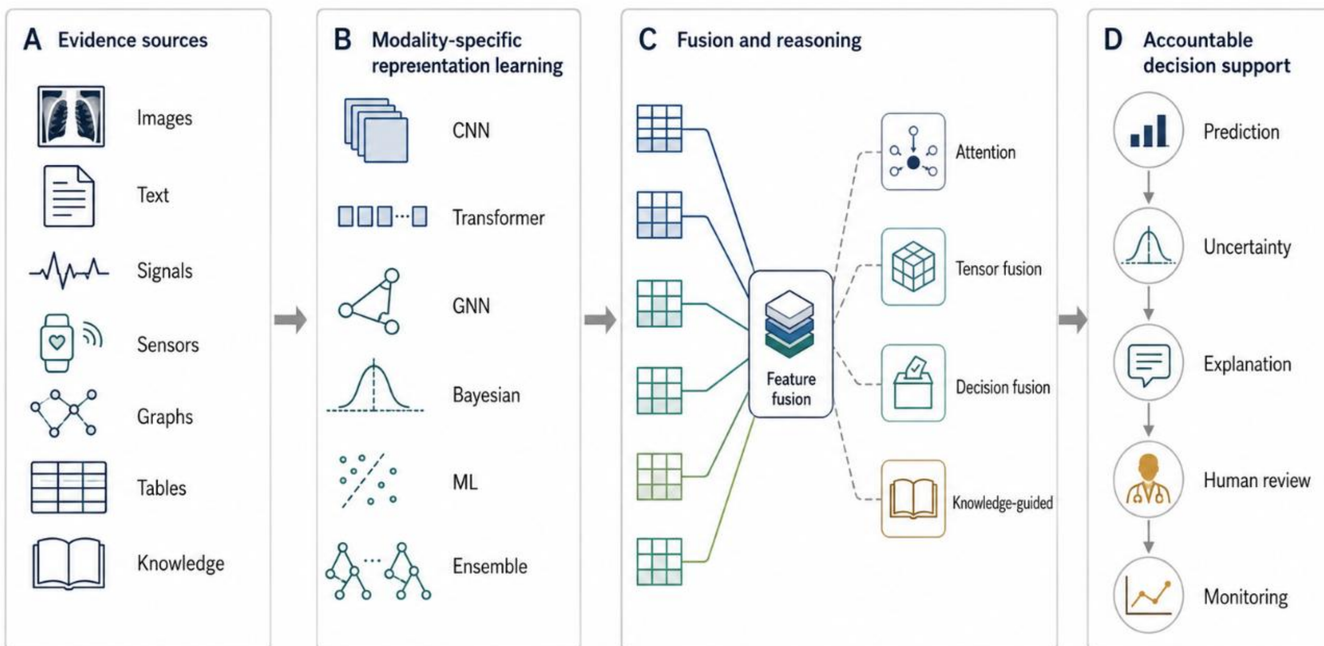

Figure 1: From heterogeneous evidence to accountable multimodal decision support.

2. Review Scope and Multimodal-Hybrid Framework

This review was assembled to span the full evidence spectrum relevant to multimodal and hybrid AI decision systems, from direct multimodal fusion studies through single-modality domain systems that motivate fusion and deployment-context papers that establish the infrastructure requirements. A seven-axis taxonomy organizes the evidence. Axis 1 classifies each paper by evidence role: direct multimodal AI evidence, hybrid architecture evidence, modality-specific evidence, domain application evidence, fusion strategy evidence, deployment-context evidence, or trustworthiness and governance evidence. Axis 2 classifies by data modality across ten categories from medical images to multimodal or multi-source data. Axis 3 characterizes the fusion level applied: data-level, feature-level, representation-level, attention-based, tensor, decision-level, ensemble or stacking, knowledge-guided, infrastructure-level, or not explicitly fusion-based but relevant as domain or modality evidence. Axis 4 identifies the architecture family across nine categories from conventional ML to edge-cloud-federated systems. Axis 5 classifies by application domain across seven sectors. Axis 6 records the decision-support function served. Axis 7 catalogues the deployment and trustworthiness concern.

This taxonomy enables two complementary forms of analysis: vertical analysis within evidence roles (characterizing what fusion strategies are used in which domains) and horizontal analysis across domains (identifying which fusion levels and trustworthiness concerns recur systematically). An important methodological point: not every paper in the corpus is explicitly multimodal or hybrid. Papers representing single-modality baselines, domain contexts, and deployment frameworks are classified as modality-specific, domain, or deployment-context evidence and used to motivate or contextualize multimodal integration rather than as direct evidence of fusion. Table 1 summarizes the seven-axis classification framework used to organize the reviewed corpus and to distinguish direct multimodal evidence from hybrid, modality-specific, deployment-context, and trustworthiness-oriented evidence.

Table 1. Seven-axis classification framework for multimodal and hybrid AI evidence.

Dimension	Core categories	Purpose
Evidence role	Direct multimodal; hybrid architecture; modality-specific; domain-context; deployment/trustworthiness evidence	Distinguishes direct fusion evidence from supporting contextual evidence
Data modality	Images; text; signals; sensors; graphs; tables; audio-visual; multimodal data	Identifies the evidence sources used for decision support
Fusion level	Data; feature; representation; attention; tensor; decision; ensemble/stacking; knowledge-guided; infrastructure-level	Classifies how heterogeneous evidence is integrated
Architecture family	ML; CNN; Transformer; CNN–Transformer; ensemble; GNN; Bayesian; generative; federated/edge-cloud AI	Maps the computational design used in each study
Application domain	Healthcare; assistive AI; industrial monitoring; IoT; agriculture; enterprise; cybersecurity	Enables cross-domain comparison of use cases
Decision-support function	Diagnosis; screening; classification; monitoring; prediction; risk assessment; recommendation; fault detection	Links model outputs to practical decision roles
Deployment/trustworthiness concern	Alignment; missingness; interpretability; robustness; calibration; privacy; security; feasibility; oversight; governance	Evaluates readiness beyond performance metrics

3. Methodological Foundations of Multimodal and Hybrid AI

3.1. Why Multimodal and Hybrid AI is Needed

Single-modality AI systems face an inherent representational ceiling: they can only learn from the information available in one data type, and no single modality captures all decision-relevant information in complex real-world settings. A structured ML model for heart disease prediction [13] uses clinical tabular records effectively, but it cannot incorporate imaging evidence from echocardiograms or temporal signal patterns from continuous monitoring without architectural extension. A CNN-based agricultural disease classifier [23] processes leaf images efficiently but cannot integrate meteorological sensor data, growth-stage records, or agronomic knowledge that experienced farmers would naturally incorporate into a diagnosis. Multimodal and hybrid architectures address this ceiling by combining complementary evidence sources, reducing vulnerability to single-modality failure, and enabling more complete representation of the decision context.

3.2. Multimodal Data and Evidence Alignment

The practical challenge of multimodal AI begins before model training: aligning heterogeneous data streams that may differ in temporal resolution, spatial scale, annotation convention, and quality as shown in Figure 2. Medical imaging data [2, 53, 75] has different preprocessing requirements from physiological signals [71, 73] or electronic health records [13]. Acoustic-emission signals [6, 16, 60] are sampled at different rates and spatial densities than visual inspection images of the same physical asset. Multimodal emotion recognition systems [14] must align facial video frames with EEG signals that capture different aspects of affective state at different temporal resolutions. Missing data is particularly consequential: if one modality is unavailable at inference time due to sensor failure, patient non-compliance, or transmission error, the fusion architecture must handle this gracefully or produce unreliable outputs. The multimodal EEG analysis framework [71] and the privacy-preserving multimodal cancer diagnosis system [1] both operate in contexts where modality availability cannot be guaranteed, making alignment and missing-modality robustness design requirements rather than optional enhancements.

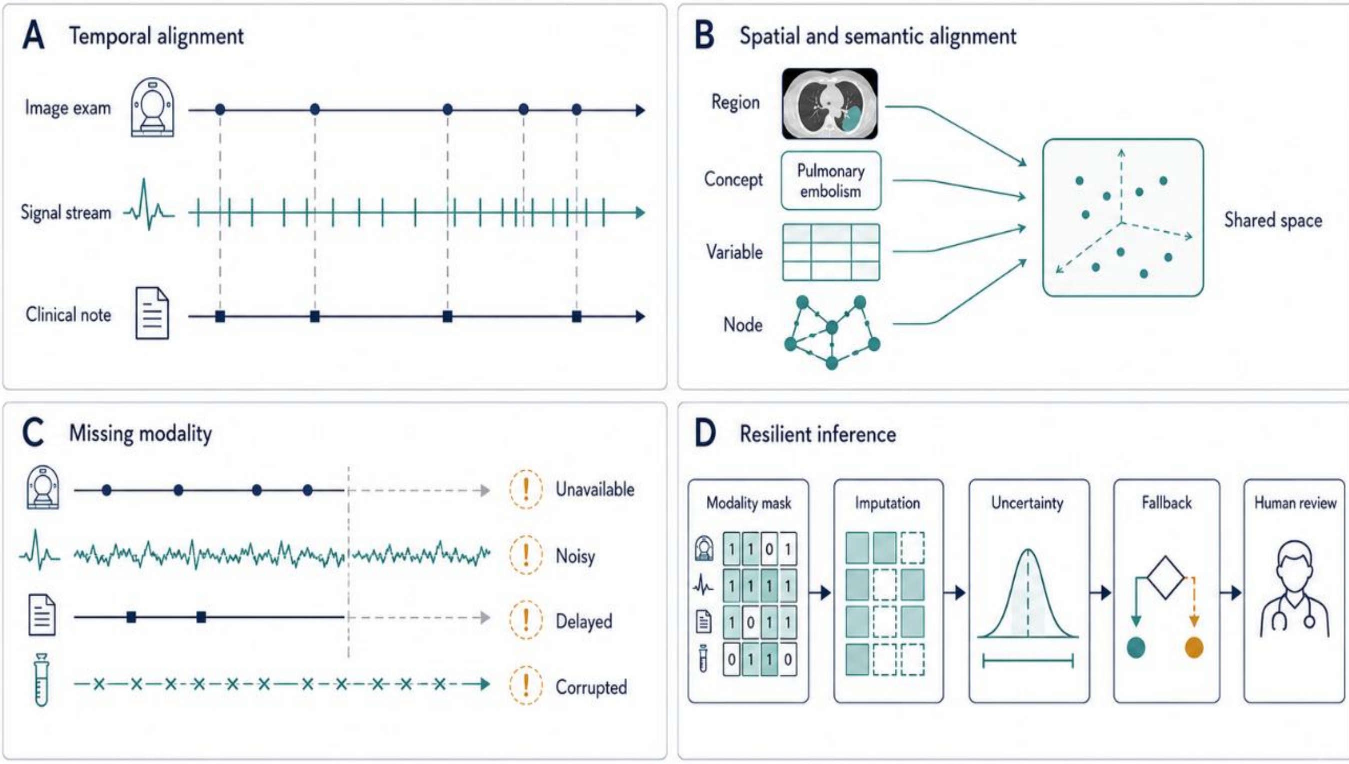

Figure 2: Alignment and missing-modality resilience in multimodal AI.

3.3. Fusion Strategies

Fusion strategy selection is among the most consequential architectural decisions in multimodal AI. Data-level fusion concatenates raw inputs before any representation learning, appropriate when modalities share a common feature space and temporal alignment is precise, but vulnerable to dimensionality explosion and missing-data cascades. Feature-level fusion combines intermediate representations learned independently from each modality, the dominant approach in medical multimodal systems [1, 14] because it allows modality-specific preprocessing and handles missing modalities by substituting mean representations. Attention-based fusion [11, 17, 21, 25, 59, 65, 68, 74, 78] dynamically weights cross-modal contributions based on contextual relevance, potentially improving adaptability but adding parameters that increase overfitting risk. Tensor fusion [48] computes outer products of modality representations, capturing cross-modal interaction terms explicitly but at quadratic computational cost. Decision-level fusion combines modality-specific model outputs, preserving modularity and enabling graceful degradation when modalities are missing, but potentially discarding inter-modal information available earlier in the pipeline. Ensemble and stacking fusion [20, 37, 43, 54, 79] aggregate predictions from heterogeneous model families, improving generalization at the cost of increased maintenance burden. Knowledge-guided fusion [24, 29, 60] integrates symbolic domain knowledge into the fusion pipeline, providing structural accountability that purely data-driven fusion cannot match.

3.4. Hybrid Architectures

Hybrid architectures combine multiple model families within a single decision pipeline, motivated by the observation that no single architecture family dominates across all representation tasks. The hybrid InceptionV3DenseNet framework for multimodal emotion recognition [14] combines inception and densely connected convolutional modules to process heterogeneous affective signals. Hybrid vision transformers for lung cancer [74] and prostate cancer [65] combine convolutional feature extraction with self-attention, inheriting the local feature sensitivity of CNNs and the long-range dependency modeling of transformers. Explainable hybrid deep learning for skin cancer [27] integrates deep feature learning with post-hoc explanation mechanisms. Stacking ensembles [37, 54, 79] combine heterogeneous base learners whose predictions are meta-learned by a second-stage model, improving generalization across training distributions. The physics-guided Bayesian neural network [62] is the most architecturally distinctive hybrid in the corpus, combining a physical domain model with a Bayesian neural network to produce uncertainty-aware predictions that neither component could generate independently.

3.5. Interpretability in Multimodal and Hybrid Systems

Interpretability challenges scale with architectural complexity. A conventional ML model [13] provides feature-level attribution that is directly auditable. A CNN with saliency maps [27, 68] provides visual evidence of attention regions that clinicians can inspect, though with known validity limitations. A stacking ensemble [37, 54] requires explanation methods that account for inter-learner interactions: explaining the meta-learner's output is not equivalent to explaining the base learners, and explanations that focus only on the final layer may misrepresent the chain of reasoning. A multimodal fusion system [1, 14, 48, 71] must explain not only what was predicted but which modality drove the prediction—a requirement that current post-hoc explanation methods do not reliably satisfy. Knowledge-graph-based fusion [24, 29] provides the most structurally auditable form of explanation: entity-linked reasoning chains that domain experts can trace and validate. The trustworthy AI framework [4] situates explainability within a broader accountability architecture that includes governance, privacy, and human oversight.

3.6. Deployment Implications

Multimodal and hybrid systems introduce deployment challenges that do not arise in single-modality models. Computational cost scales with the number of modalities and model components, directly affecting inference latency. Web-based deployment [11, 79] and IoT-embedded monitoring [42, 49, 63] impose strict latency budgets that may require model compression, modality selection, or cloud offloading. Privacy-preserving deployment [1, 9, 15] must protect each modality independently—a federated learning framework for multimodal clinical data must maintain differential privacy guarantees across imaging, text, and physiological modalities simultaneously. Maintenance complexity increases with the number of independently trainable components: when one modality's sensor drift causes performance degradation, root-cause attribution in a stacked or fused pipeline is substantially harder than in a single-modality system.

4. Architecture Families for Multimodal and Hybrid Decision Systems

4.1. Conventional Machine Learning and Structured Analytics

Conventional machine learning constitutes the most deployment-accessible architecture family for structured, tabular, or text-based decision support, and provides the modality-specific baselines against which hybrid systems are measured. Clinical decision support for heart disease using structured patient data [13] and personalized ML for Parkinson's screening via voice biomarkers [73] illustrate the range of structured and physiological ML applications. In business analytics, market trend forecasting with external factor integration [51], retail demand forecasting using LSTM and gradient boosting [19], e-commerce pricing optimization [18], and small-business management ML [58] demonstrate that conventional ML architectures serve as operational baselines for enterprise decision support. Multi-class Bengali social media sentiment classification [8] and drug review sentiment extraction [5] provide text-modality evidence. The deployment relevance of conventional ML as a fusion component is significant: structured ML models can serve as decision-level fusion aggregators for multi-source inputs, or as modality-specific modules whose predictions are combined by higher-level ensemble or meta-learning stages.

4.2. CNN-Based Deep Learning and Transfer Learning

CNN-based architectures provide the dominant modality-specific representation-learning foundation for image, signal, and acoustic-emission-based AI decision systems. Transfer learning for sleep stage classification under data-constrained conditions [30] and early leukemia diagnostics using image processing and transfer learning [75] illustrate how pre-trained CNN features accelerate domain-specific representation learning. Facial emotion recognition via a bidirectional Elman neural network [67] and a hybrid deep belief optimization system [28] extend CNN-based representation learning to affective computing. The lightweight deep learning system for concrete crack characterization via acoustic-emission signals [6] demonstrates CNN compressibility for edge deployment in industrial monitoring. Multichannel CNN analysis of imbalanced CT data for lung cancer [2] illustrates how multichannel inputs provide a structured pathway toward multimodal integration within a convolutional framework. Lightweight ResNeXt for aquaculture disease [31] and advanced deep learning for tea leaf disease [32] provide agricultural modality evidence. Iris detection and recognition [44] extends CNN-based systems to biometric identification. In breast cancer diagnosis, neural network–based approaches combined with dimensionality reduction have shown potential for extracting discriminative patterns from complex morphological data, while optimized neural architectures further enhance classification reliability and computational efficiency [80], [81]. Beyond diagnosis, AI also plays a critical role in securing healthcare and essential infrastructure by enabling intelligent threat detection, risk monitoring, and adaptive cybersecurity defense mechanisms, which are necessary for protecting sensitive medical systems and ensuring dependable digital health operations [82].

4.3. Vision Transformers and Attention-Based Models

Vision transformers have become the dominant architecture for image-based medical and agricultural classification, with self-attention mechanisms providing both representational flexibility and explainability potential. The Swin Transformer for cervical cell classification with XAI and web deployment [11], the MaxViT soybean disease model [23], the MaizeFormerX lightweight cross-scale ViT [59], the global-local attention model for kidney disease classification [25], FuseAttenX attention-enhanced deep learning for business strategy optimization [21], and the explainable transformer for skin lesion classification [68] each illustrate distinct attention-based architectural configurations. The dual-branch visual transformation models for ASD classification [17, 41] demonstrate how multi-branch attention architectures can combine different feature perspectives within a single forward pass, an architectural pattern directly applicable to multi-source fusion. The critical observation for multimodal systems is that attention maps within a unimodal transformer show which input regions influenced the prediction, but cross-modal attention, how one modality modulates attention in another, requires explicit architectural design and should not be assumed from unimodal attention mechanisms.

4.4. Ensemble, Stacking, and Hybrid Deep Learning Systems

Ensemble and stacking architectures represent the most widely used hybrid strategy in the corpus, combining multiple base learners to reduce variance, improve generalization, and enable richer post-hoc explanation. The explainable deep stacking ensemble for brain tumor diagnosis [37], the hierarchical Swin Transformer ensemble for breast cancer with decentralized deployment [20], the stacking ensemble for breast cancer with real-time web deployment [79], the hybrid vision transformer for prostate cancer in MRI [65], the stacking ensemble for cervical cancer with XAI [54], and the ViX-MangoEFormer ensemble for mango disease recognition with XAI [43] collectively demonstrate the versatility of ensemble and stacking fusion across oncological imaging and agricultural disease detection. The ensemble transformer with post-hoc XAI for depression emotion and severity detection [3] illustrates the extension of this approach to affective computing. The explainable AI hybrid deep learning framework for skin cancer [27] and the LMVT hybrid vision transformer for lung cancer [74] integrate convolutional and transformer components. The explainable transformer for cotton leaf diagnostics and fabric defect detection [78] extends hybrid architectures to dual-task agricultural and industrial inspection.

4.5. Multimodal Fusion Systems

Direct multimodal fusion systems—those explicitly designed to process and integrate two or more distinct data modalities—represent the core evidence category for this review. The multimodal machine learning framework for privacy-preserving and scalable cancer diagnosis [1] addresses clinical AI at the intersection of multimodal integration and privacy-preserving deployment, a combination that reflects the operational demands of real-world healthcare systems. The hybrid multi-modal emotion recognition framework based on InceptionV3DenseNet [14] explicitly combines EEG signals and visual inputs to recognize emotion states more comprehensively than either modality alone. The vision-audio multimodal object recognition system using hybrid and tensor fusion [48] is the most architecturally explicit multimodal fusion system in the corpus: tensor fusion computes explicit cross-modal interaction terms between visual and auditory representations, enabling the model to capture correlations that feature concatenation would suppress. The multimodal EEG analysis framework for neural synchrony in phrase processing [71] integrates multiple EEG channels with contextual linguistic information, illustrating multimodal fusion in neuroscientific decision support. Each of these systems illustrates a distinct fusion strategy privacy-preserving infrastructure-level integration, feature-level EEG+visual fusion, tensor product fusion, and multichannel EEG integration that addresses a different aspect of the multimodal design space.

4.6. Graph Neural Networks and Knowledge-Graph Reasoning

Graph neural networks and knowledge-graph architectures provide a form of relational fusion that is structurally distinct from statistical feature combination. The GNN-enhanced acoustic-emission gas-pipeline monitoring system [60] models signal propagation across sensor network topology, integrating spatial relational structure into fault diagnosis that purely signal-level classifiers cannot access. Knowledge-graph and NLP integration for heuristic reasoning [24] and the AddManBERT knowledge-graph construction for additive manufacturing design support [29] demonstrate that symbolic knowledge graphs can be fused with neural language representations to produce reasoning chains that are both expressive and auditable. From a fusion perspective, knowledge-guided fusion is particularly valuable in domains where expert knowledge is stable and explainability is required by professional standards—engineering, clinical decision support, and regulatory compliance contexts where black-box feature fusion is insufficient.

4.7. Bayesian, Physics-Guided, and Uncertainty-Aware Systems

The physics-guided Bayesian neural network for sensor fault detection in wind turbines [62] represents a fusion strategy that combines a physical domain model encoding known relationships between sensor measurements and turbine states with a Bayesian neural network that quantifies epistemic uncertainty over sensor fault classifications. This architecture is a knowledge-guided fusion system in the fullest sense: the physical model provides structural prior knowledge that constrains the Bayesian network's hypothesis space, and the Bayesian framework produces calibrated uncertainty estimates that can trigger human oversight when the model's confidence is insufficient for autonomous action. The deployment significance of this architecture extends beyond wind turbines: any safety-critical multimodal monitoring system that processes heterogeneous sensor streams in the presence of sensor noise, degradation, or novel fault patterns requires uncertainty-aware inference of this kind.

4.8. Generative, Enterprise, Distributed, and Privacy-Preserving AI

Infrastructure-level integration, the deployment of AI within federated, edge-cloud, enterprise information system, and cybersecurity frameworks—constitutes the operational layer that makes multimodal AI systems viable at scale. The distributed edge-cloud-6G federated learning framework for secure and auditable decision support [15] provides the architectural backbone for privacy-preserving multimodal learning across institutions. Privacy-preserving behavior analytics for workforce retention [9] demonstrates operational differential privacy in organizational analytics. Generative AI in enterprise information systems [66] addresses the strategic layer of enterprise AI where generated recommendations draw on heterogeneous organizational data. The intelligent cybersecurity framework for ML-driven data protection and threat intelligence [64] illustrates multi-source security analytics where threat evidence is inherently heterogeneous. AI as a strategic engine for data security and digital communication resilience [47] and the resilience-by-design framework [38] provide the governance and systemic resilience layer that underpins all infrastructure-level multimodal integration.

5. Cross-Domain Application Synthesis

Table 2. Cross-domain multimodal AI application landscape.

Domain	Evidence sources	Decision role	Opportunity	Deployment gap
Healthcare	Images, EHR, pathology, signals, text	Diagnosis, screening, prognosis	Combine clinical and imaging evidence	Validation, calibration, privacy
Assistive / affective AI	EEG, face, speech, text, behavior	Emotion, ASD, mental-health support	Fuse human behavioral signals	Alignment, fairness, user variability
Industrial systems	Acoustic, vibration, thermal, images, topology	Fault detection, monitoring	Link sensor data with physical context	Drift, missing sensors, latency
IoT / infrastructure	IoT streams, meters, network data	Monitoring, routing, optimization	Distributed real-time intelligence	Edge feasibility, packet loss
Agriculture	Plant images, weather, soil, sensors	Disease detection, yield planning	Add field context to image models	Field validation, sensor alignment
Enterprise analytics	Transactions, market data, workflows, text	Forecasting, pricing, risk, strategy	Integrate operational and strategic data	Governance, bias, auditability
Cybersecurity	Logs, traffic, behavior, threat intelligence	Threat detection, secure decisions	Combine cyber and behavioral evidence	Adversarial risk, privacy, coordination

5.1. Healthcare and Biomedical Decision-Making

Healthcare is the domain where multimodal and hybrid AI has the greatest potential impact and the most stringent trustworthiness requirements. The direct multimodal evidence is provided by the privacy-preserving multimodal cancer diagnosis framework [1], which integrates multiple clinical data sources under privacy constraints. Across oncological imaging, the corpus spans skin cancer [27, 68], lung cancer [2, 74], breast cancer [20, 79], cervical cancer [11, 54], brain tumor [37], leukemia [75], prostate cancer [65], cytological cancer classification [53], and kidney disease classification [25]. Each of these applications provides modality-specific or hybrid architecture evidence that motivates the broader multimodal integration agenda: a system that diagnoses cancer from a single image modality is less informationally complete than one that integrates imaging with pathological, genomic, or clinical record evidence. Heart disease prediction from structured data [13], diabetes management through AI-integrated health information systems [10], sleep stage classification [30], and Parkinson's screening via voice biomarkers [73] illustrate the physiological and structured data modalities available for clinical multimodal fusion. The ensemble transformer with post-hoc XAI for depression detection [3] extends multimodal evidence to mental health, where affective and textual signals are natural fusion candidates. Web-based screening deployment [11, 79] demonstrates that multimodal decision support can be packaged for remote access, provided latency and privacy requirements are met. Table II summarizes the cross-domain application landscape of multimodal and hybrid AI, highlighting the dominant evidence sources, decision-support functions, architectural patterns, multimodal opportunities, and deployment gaps across the seven application sectors reviewed in this study.

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

The human-centered and assistive AI domain is where multimodal evidence is most directly demanded by the complexity of human behavior. The hybrid multimodal emotion recognition framework [14] and the multimodal EEG neural synchrony analysis [71] represent direct multimodal fusion of physiological and perceptual signals for affective computing. ASD classification using dual-branch visual transformation [17, 41] demonstrates attention-based architectural fusion for facial and emotion-feature evidence. The ASD facial expression database [45] provides the foundational modality resource. The AI-powered digital health platform for ASD students [12] illustrates the deployment context where adaptive, personalized, and therapeutically accessible AI is needed. Facial emotion recognition systems [28, 67] and the hybrid InceptionV3DenseNet framework [14] span multiple CNN evolutionary stages in affective signal processing. Suicidal ideation detection using NLP [22] and Bengali social media sentiment classification [8] provide text-modality evidence for mental health and social computing. Drug review sentiment extraction [5] bridges health information and text analytics. The adaptive feedback system for learner improvement [26], the flex sensor hand glove for deaf and mute individuals [39], iris detection and recognition [44], and the tDCS clinical model [61] extend the assistive AI context to education, physical accessibility, biometrics, and neuromodulation.

5.3. Industrial Monitoring, Cyber-Physical Systems, and Robotics

Industrial monitoring AI is characterized by the co-presence of multiple sensor modalities—acoustic emission, vibration, thermal, visual, and process parameter streams that individually provide partial evidence of equipment state and jointly provide richer diagnostic coverage. The GNN-enhanced gas-pipeline monitoring system [60] and the multivariate acoustic-emission imaging system for gas-pipeline diagnosis [16] represent two evolutionary stages, statistical imaging and graph-relational reasoning, in the same pipeline monitoring domain. The lightweight deep learning system for concrete crack characterization via acoustic-emission signals [6] provides edge-deployable modality evidence for structural health monitoring. The physics-guided Bayesian neural network for wind-turbine sensor fault detection [62] provides the most principled fusion of physical knowledge and probabilistic inference in the industrial cluster. The vision-audio multimodal object recognition system via tensor fusion [48] contributes direct multimodal evidence for industrial perception in robotics and quality control. The question of full autonomy in underwater robotics [56] engages the human oversight axis of industrial agentic AI—a governance question with direct implications for how much independent action multimodal monitoring AI should take in safety-critical environments.

5.4. Smart Infrastructure, IoT, Energy, and Communication Systems

Smart infrastructure AI operates in intrinsically multi-source environments: IoT-based solar micro-grid monitoring [42], smart energy metering [63], and smart healthcare medical boxes for elderly patients [49] each aggregate data from multiple sensor types whose combined analysis provides infrastructure oversight that no single sensor can achieve. Wireless mesh network routing optimization [35] and MANET routing protocol simulation [72] address network-layer decision support in distributed infrastructure. High-altitude platform communications optimization [57] extends infrastructure AI to airborne communication systems. The multi-source IoT evidence in this domain motivates infrastructure-level fusion architectures: not feature-level or attention-based fusion of pre-defined modality pairs, but deployment-aware integration of heterogeneous IoT streams with varying quality, latency, and availability profiles. The distributed intelligence and edge-cloud-6G federated learning framework [15] provides the deployment architecture for privacy-preserving multimodal integration at infrastructure scale.

5.5. Agriculture, Environment, and Sustainability

Agricultural AI illustrates the evolution from task-specific single-modality image classifiers toward lightweight, explainable, and field-deployable systems where image and environmental sensor fusion is increasingly recognized as the next step. MaizeFormerX lightweight cross-scale ViT [59], MaxViT soybean disease model [23], ViX-MangoEFormer ensemble with XAI [43], explainable transformer for cotton leaf diagnostics [78], advanced deep learning for tea leaf disease [32], and lightweight ResNeXt for aquaculture disease [31] collectively constitute the precision agriculture disease detection cluster. Each currently operates in a single image modality but provides the foundational classifiers that a multimodal system, integrating image evidence with weather data, soil sensor readings, and growth-stage records, would use as modality-specific modules. AI-driven smart agriculture for crop yield optimization [7] and AI-driven solar financing for rural health businesses [69] address the systemic sustainability dimension. The resilience-by-design framework [38] provides the cross-sectoral lens connecting agricultural sustainability, rural health, and infrastructure resilience.

5.6. Business, Enterprise, and Organizational Decision-Making

Business and enterprise AI draws on the widest diversity of data modalities in the corpus—structured transactional records, social media text, market indicators, blockchain logs, workflow data, and organizational knowledge—making it a natural target for multimodal integration even though most current systems process one modality at a time. Credit scoring for financially underserved businesses [52] and predictive project risk analytics [33] illustrate the tabular modality. Market basket analysis for healthcare bundling [34] and customer satisfaction analytics in hospitality [36] illustrate cross-tabular analytical AI. Blockchain and ML in supply chain management [40] introduces distributed ledger data as a trust-structured modality. AI for agile IT project risk and decision-making [70], automated risk assessment AI in agile project management [50], and digital transformation analytics [46] address governance-structured business contexts. Generative AI in enterprise information systems [66], AI-driven business analytics for IT strategy [77], AI-enabled MIS for economic resilience and governance [55], and AI-ERP integration in dark factories [76] address the strategic and enterprise integration layer where multi-source AI is operationally required. The attention-enhanced deep learning system FuseAttenX for business strategy optimization [21] illustrates hybrid attention-based analytics in enterprise decision support. Small-business management ML [58] completes the business modality coverage.

5.7. Cybersecurity, Privacy, and Distributed Intelligence

Cybersecurity AI is inherently multimodal in its evidence requirements: effective threat detection draws simultaneously on network traffic patterns, user behavior analytics, log data, and external threat intelligence, modalities with different temporal dynamics, formats, and availability profiles. The intelligent cybersecurity ML framework for data protection and threat intelligence [64] addresses this multi-source security analytics context. Privacy-preserving behavior analytics for workforce retention [9] demonstrates operational privacy-preserving organizational analytics. AI as a strategic engine for data security and digital communication resilience [47] positions security AI at the organizational governance level. The distributed edge-cloud-6G federated learning framework [15] provides the infrastructure for privacy-preserving multimodal learning across distributed nodes. Trustworthy AI for high-stakes decision support across critical sectors [4] and the resilience-by-design framework [38] provide the cross-sector governance architecture within which all cybersecurity multimodal AI must operate.

6. Challenges in Multimodal and Hybrid AI for Real-World Decision-Making

6.1. Modality Alignment and Missing Data

The practical challenge of multimodal AI begins with alignment: heterogeneous data streams that differ in temporal resolution, spatial scale, annotation convention, and quality must be synchronized before fusion. Medical imaging and physiological signals [71, 73] require temporal alignment that clinical workflows may not consistently provide. IoT sensor streams [42, 49, 63] have variable sampling rates and packet loss profiles. When one modality is unavailable at inference time, a sensor failure, patient non-compliance, or communication outage, a fusion architecture that cannot degrade gracefully produces unreliable outputs. Missing-modality robustness must be evaluated through systematic modality-dropout testing, which remains inconsistently reported across the corpus.

6.2. Fusion Design and Overfitting

Complex fusion architectures increase the number of learnable parameters, raising the risk of overfitting, particularly when training datasets are small relative to model capacity—a common condition in medical and industrial AI. Tensor fusion [48] computes cross-modal outer products, expanding parameter count quadratically with modality representation dimensionality. Stacking ensembles [37, 54, 79] introduce a second training stage whose generalization depends on base learner diversity that may not be achievable with limited data. Attention-based fusion [11, 25, 65, 74] adds cross-attention parameters that are poorly constrained without large-scale multimodal pretraining. Fusion ablation, systematically removing each modality or fusion component and evaluating the resulting performance, is the standard method for assessing fusion contribution but is inconsistently reported in the literature.

6.3. Interpretability and Explanation Conflict

When multiple modalities produce explanations that conflict, one modality's saliency map highlights a region that another modality's feature attribution deemphasizes, the fusion system's output explanation becomes ambiguous. Post-hoc attribution methods applied to the fused output [3, 27, 37] cannot distinguish whether the prediction was driven by one modality, the other, or their interaction. Knowledge-graph reasoning [24, 29, 60] provides the most conflict-resistant form of explanation by explicitly representing causal and relational pathways but requires curation effort that scales with knowledge domain complexity. The trustworthy AI framework [4] addresses explanation accountability at the system level, but domain-specific explanation validation protocols for multimodal clinical and industrial systems remain underdeveloped.

6.4. Robustness and Distribution Shift

Distribution shift in multimodal systems can originate from any single modality, and its propagation through the fusion pipeline may produce correlated failures that are harder to detect than single-modality degradation. Medical imaging cross-scanner shifts [1, 20] may co-occur with electronic health record coding changes, creating compound distribution shifts that no single-modality robustness evaluation could detect. Industrial sensor degradation [62] may produce signals that fall outside the physics-guided Bayesian model's prior distribution, requiring uncertainty estimates that trigger human review rather than autonomous action. Business forecasting models [19, 51] face economic regime changes that affect multiple tabular modalities simultaneously. Robustness evaluation in multimodal systems requires modality-specific distribution shift testing as well as compound-shift scenarios.

6.5. Privacy, Security, and Federated Integration

Privacy-preserving multimodal AI requires protecting each modality independently while maintaining the utility of cross-modal fusion. The multimodal privacy-preserving cancer diagnosis framework [1] addresses this at the system level. The federated learning framework [15] enables collaborative multimodal learning across data-sovereign institutions. Privacy-preserving workforce analytics [9] demonstrates differential privacy in a single-modality organizational context; extending this to multimodal employee monitoring data introduces additional privacy accounting complexity. The cybersecurity layer of multimodal systems [64, 47] must account for adversarial attacks targeting specific modality inputs—adversarial perturbations designed to mislead one modality while leaving others unchanged may still corrupt fused predictions through the shared fusion layer.

6.6. Computational Complexity and Deployment Feasibility

Multimodal inference latency is additive across modalities and multiplicative across fusion interaction terms. Tensor fusion [48], while capturing cross-modal interactions explicitly, may be computationally infeasible for real-time IoT deployment [49, 63] or edge medical devices. Lightweight single-modality architectures [6, 31, 59, 78] demonstrate that model compression is achievable for edge deployment in agricultural and industrial contexts; extending this to multimodal systems without disproportionate performance loss is an active research challenge. Web-based deployment of multimodal diagnostic systems [11, 79] requires careful pipeline engineering to manage inference latency under variable network conditions and device capabilities.

6.7. Human Oversight, Governance, and Accountability

Multimodal and hybrid AI systems should, by default, be designed to support human decision-making rather than replace it. The question of full autonomy in underwater robotics [56] illustrates the governance challenge at the autonomous frontier: when a multimodal system operating in an unstructured environment takes consequential actions without human review, accountability must be explicitly assigned and documented. The trustworthy AI framework [4] and the resilience-by-design framework [38] address governance at the system level. Generative AI systems [66] that integrate multi-source enterprise data to generate strategic recommendations introduce hallucination risk that is particularly difficult to detect in multimodal contexts, where a plausible-sounding output may reflect spurious cross-modal correlations rather than genuine insight.

6.8. Evidence Maturity and Reproducibility

Evidence maturity in multimodal AI requires not only standard performance reporting but fusion-specific validation: ablation studies demonstrating each modality's contribution, modality-dropout testing characterizing graceful degradation, calibration analysis showing that fused confidence scores are reliable, and reproducibility documentation sufficient to replicate the fusion pipeline. The comparative explainable ML analysis for cancer cytology [53] and the personalized Parkinson's screening model [73] illustrate systematic comparative evaluation designs for single-modality systems; analogous designs for multimodal systems would include modality contribution comparisons. The current corpus does not consistently provide fusion ablation evidence, indicating that maturity standards for multimodal and hybrid AI remain below those expected in clinical or industrial deployment contexts.

7. Future Research Directions

Future research should prioritize standardized benchmarks, rigorous validation, and governance-aware deployment. Cross-domain multimodal benchmarks are needed with aligned modalities, missing-modality splits, and distribution-shift scenarios, reporting unimodal and fused performance, modality-dropout accuracy, and robustness [1,4]. Studies should also include systematic fusion ablation to quantify each modality's marginal contribution, ablation profile, and cross-modal interaction strength [14,48]. Because real-world inputs may be missing, noisy, delayed, or corrupted, future models should evaluate k-of-n modality availability, graceful degradation, and calibrated alert thresholds [62,71]. Explanation methods must also move beyond attention visualization by assessing fidelity, cross-modal agreement, and expert/user comprehension in hybrid and stacking systems [4,27,37]. Privacy-preserving multimodal learning should extend differential privacy, secure aggregation, and federated learning to clinical and organizational settings while reporting modality-specific privacy budgets, federated utility loss, and communication efficiency [1,9,15]. Efficient federated and edge-deployable fusion architectures are also required for bandwidth- and hardware-constrained environments, with evaluation of latency, memory footprint, and accuracy–efficiency trade-offs [6,15,59]. Lightweight hybrid systems remain especially important for IoT, agriculture, and point-of-care applications, where modality-wise parameter burden and inference cost should be reported [6,31,59,78]. Graph-enhanced and knowledge-guided fusion should be advanced to support auditable reasoning chains, evaluated through knowledge-graph coverage, reasoning fidelity, and expert assessment [24,29,60]. Human-in-the-loop protocols should compare decision quality with and without AI support, override frequency, and modality-specific trust calibration [4,56]. Finally, governance-aware reporting standards should document modalities, alignment, fusion design, ablation, missing-modality handling, validation, and safeguards [4,38], while evidence-maturity levels should distinguish proof-of-concept fusion from externally validated and deployment-ready multimodal AI.

8. Limitations of the review

The synthesis is thematic, methodological, fusion-oriented, and deployment-level rather than quantitative. Specific performance metrics, fusion ablation results, modality alignment strategies, dataset characteristics, validation protocols, computational requirements, deployment environments, user studies, and statistical evidence could not be extracted from titles alone. 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, experimental details, and supplementary materials would be required to support meta-analytic comparisons of fusion strategies, modality contribution, explanation quality, or deployment feasibility. The curated corpus may not comprehensively represent all multimodal AI research threads; video-based fusion, genomic-imaging integration, multi-agent multimodal systems, and natural language–image grounding are underrepresented. The seven-axis taxonomy is one defensible organization; alternative taxonomies emphasizing different fusion properties may yield complementary insights.

9. Conclusion

This structured critical review has examined multimodal and hybrid AI for real-world decision-making across seven application domains, healthcare and biomedical AI, human-centered and assistive AI, industrial monitoring and cyber-physical systems, smart infrastructure and IoT, agriculture and sustainability, business and enterprise analytics, and cybersecurity and distributed intelligence, using a seven-axis taxonomy classifying papers by evidence role, modality, fusion level, architecture family, domain, decision-support function, and deployment or trustworthiness concern. The synthesis reveals a field in which fusion strategies have diversified from early feature concatenation through attention-based, tensor, stacking, ensemble, and knowledge-guided fusion, with direct multimodal evidence in affective computing, medical diagnosis, industrial perception, and privacy-preserving clinical AI. Hybrid architecture combining CNNs, transformers, graph neural networks, Bayesian models, and ensemble systems, provide the representational diversity needed to address the heterogeneity of real-world evidence, while the broader corpus of single-modality and domain papers establishes the application contexts and baseline modalities that multimodal systems must surpass to justify their additional complexity.

The path toward trustworthy multimodal and hybrid AI in real-world deployment requires addressing the validation gaps identified throughout this review: fusion ablation to verify modality contribution, modality-dropout testing to ensure graceful degradation, explanation validation to confirm that multimodal explanations faithfully represent cross-modal reasoning, calibrated uncertainty reporting to support human oversight decisions, and governance-aware reporting standards that make fusion design choices transparent to practitioners and regulators. Privacy-preserving multimodal learning federated and edge-deployable fusion architectures, and knowledge-guided integration that provides structural auditability are the architectural frontiers on which progress is most urgently needed. AI systems that integrate heterogeneous evidence reliably, explain their fusion reasoning credibly, protect sensitive data rigorously, and support human oversight consistently will be the foundation of trustworthy real-world decision-making across all critical domains.

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] Kabir AA, Mahmud FU, Rahman MS, 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;1(6).
[2] Sohaib M, Hasan MJ, Zheng Z. A multichannel analysis of imbalanced computed tomography data for lung cancer classification. Measurement Science and Technology. 2024;35(8):085401. doi:10.1088/1361-6501/ad437f.
[3] 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.
[4] 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;18(3). doi:10.30574/wjaets.2026.18.3.0152.
[5] Haque R, Laskar SH, Khushbu KG, Hasan MJ, Uddin J. Data-driven solution to identify sentiments from online drug reviews. Computers. 2023;12(4):87. doi:10.3390/computers12040087.
[6] Habib MA, Hasan MJ, Kim JM. A lightweight deep learning-based approach for concrete crack characterization using acoustic emission signals. IEEE Access. 2021;9:104029-104050.
[7] Riipa MB, Saha S, Ferdousmou J, Khatoon R, Mohammad N, Hossain M. AI-driven smart agriculture: Optimizing crop yield and sustainability in the U.S. In: 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET); 2025; Paris, France. doi:10.1109/ICECET63943.2025.11472088.
[8] 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;4:21-35. doi:10.1016/j.ijcce.2023.01.001.
[9] 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.
[10] 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.
[11] 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;7(5):25-35. doi:10.32996/jmhs.2026.7.5.5.
[12] 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.
[13] Rashid SU, Siddiqui MIH, Mahmud FU, Rahman MS, Kabir AA, et al. Machine learning based clinical decision support for heart disease prediction using structured patient data. Journal of Computer Science and Technology Studies. 2024;6(1). doi:10.32996/jcsts.2024.6.1.36.
[14] Alamgir FM, Alam MS. Hybrid multi-modal emotion recognition framework based on InceptionV3DenseNet. Multimedia Tools and Applications. 2023;82(26):40375-40402. doi:10.1007/s11042-023-15066-w.
[15] 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;18(3):268-279. doi:10.30574/wjaets.2026.18.3.0154.
[16] 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. doi:10.1080/10589759.2025.2456088.
[17] Alamgir FM, Zaman T. Classification model for autism spectrum disorder individuals: Utilizing facial grid-wise emotion features and dual-branch visual transformation. In: 2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications (PEEIACON). IEEE; 2024. doi:10.1109/PEEIACON63629.2024.10800506.
[18] Chowdhury MS, Shak MS, Devi S, Miah MR, Al Mamun A, Ahmed E, Hera SAA, Mahmud F, Mozumder MSA. Optimizing e-commerce pricing strategies: A comparative analysis of machine learning models for predicting customer satisfaction. The American Journal of Engineering and Technology. 2024;6(09):6-17. doi:10.37547/tajet/Volume06Issue09-02.
[19] Shak MS, Mozumder MSA, Hasan MA, Das AC, Miah MR, Akter S, Hossain MN. Optimizing retail demand forecasting: A performance evaluation of machine learning models including LSTM and gradient boosting. The American Journal of Engineering and Technology. 2024;6(09):67-80. doi:10.37547/tajet/Volume06Issue09-09.
[20] Ahmed MR, Rahman H, Limon ZH, Siddiqui MIH, Khan MA, Pranta ASUK, Haque R, Swapno SMMR, Cho YI, Abdallah MS. Hierarchical Swin transformer ensemble with explainable AI for robust and decentralized breast cancer diagnosis. Bioengineering. 2025;12(6):651. doi:10.3390/bioengineering12060651.

[21] Mahmud FU, Rahman A, Khan MA, Bishnu KK, Eva AA, Maua J. FuseAttenX: Leveraging attention-enhanced deep learning for business strategy optimization. In: 2025 IEEE 4th International Conference on Computing and Machine Intelligence (ICMI). 2025. doi:10.1109/ICMI65310.2025.11141140.

[22] Haque R, Islam N, Islam M, Ahsan MM. A comparative analysis on suicidal ideation detection using NLP, machine, and deep learning. Technologies. 2022;10(3):57. doi:10.3390/technologies10030057.

[23] Pranta ASUK, Fardin H, Debnath J, Hossain A, Sakib AH, Ahmed MR, Haque R, Reza AW, Dewan MAA. A novel MaxViT model for accelerated and precise soybean leaf and seed disease identification. Computers. 2025;14(5):197. doi:10.3390/computers14050197.

[24] Haruna A, Noman K, Li Y, Makanda ILD, Zubair A, Hasan MJ, Alhassan AB. Facilitating heuristic reasoning by utilizing knowledge graph and natural language processing. Knowledge-Based Systems. 2026;334:115153. doi:10.1016/j.knosys.2025.115153.

[25] 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;7(5):36-45. doi:10.32996/jmhs.2026.7.5.6.

[26] 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.

[27] Al Sakib A, Swapno SMMR, Ahamed F, Mohiuddin AB, Bhuiyan MIH, Khan S, Khushbu KG, Haque R, Alahmadi TJ, Moni MA. Explainable AI-driven hybrid deep learning framework for accurate skin cancer diagnosis. Digital Health. 2026;12:20552076261438923. doi:10.1177/20552076261438923.

[28] Alamgir FM, Alam MS. An artificial intelligence driven facial emotion recognition system using hybrid deep belief rain optimization. Multimedia Tools and Applications. 2023;82:2437-2464. doi:10.1007/s11042-022-13378-x.

[29] 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;67:103578. doi:10.1016/j.aei.2025.103578.

[30] Mahmud FU, Rahman H, Limon ZH, Khan MA, Jashim FB. Transfer learning approach for sleep stage classification with limited training data. International Journal of Science and Research Archive. 2025;15(2). doi:10.30574/ijsra.2025.15.2.1506.

[31] 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.

[32] 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. In: 2025 IEEE International Conference on Emerging Technologies and Applications (MPSec ICETA); 2025. doi:10.1109/MPSecICETA64837.2025.11118779.

[33] Tanim SH, Ahmad MS, Mithun MMU, Tarannum R, Refat FR, Sunny MNM. Leveraging predictive analytics for risk identification and mitigation in project management. Journal of Information Systems Engineering and Management. 2025;10(43s):1041-1052. doi:10.52783/jisem.v10i43s.8523.

[34] Rimon RH, Nurujjaman, Mithun MMU. Market basket analysis for healthcare services to identify bundled care offerings. Frontiers in Computer Science and Artificial Intelligence. 2025;4(3):44-67.

[35] Alamgir FM, Ahmed F, Miah M, Munna HM, Barua S. A novel routing algorithm for inter-group load balancing in wireless mesh networks. In: 2018 21st Saudi Computer Society National Computer Conference (NCC). IEEE; 2018. doi:10.1109/NCG.2018.8593192.

[36] Talukder T, Masud SB, Miah MR, Hera A, Faruque MO. 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;12:1-15. doi:10.4236/oalib.1112802.

[37] Haque R, Khan MA, Rahman H, Khan S, Siddiqui MIH, Limon ZH, Swapno SMMR, Appaji A. Explainable deep stacking ensemble model for accurate and transparent brain tumor diagnosis. Computers in Biology and Medicine. 2025;191:110166. doi:10.1016/j.compbiomed.2025.110166.

[38] Shakil MR, Hasan M, Tarek MIH, 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;18(3):254-267. doi:10.30574/wjaets.2026.18.3.0153.

[39] Al Mamun A, Polash MSJK, Alamgir FM. Flex sensor based hand glove for deaf and mute people. International Journal of Computer Networks and Communications Security. 2017;5(2):38-48.

[40] Rahman T, Uddin MK, Hosen MM, Bhattacharjee B, Taluckder MS, Mou SN, Akter P, Hossain MS, Miah MR, Rahman MM. Blockchain applications in business operations and supply chain management by machine learning. International Journal of Computer Science & Information System. 2024;9(11):17-30. doi:10.55640/ijcsis/Volume09Issue11-03.

[41] 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. Biomedical Signal Processing and Control. 2026;120:109999. doi:10.1016/j.bspc.2026.109999.

[42] 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.

[43] Noman AA, et al. ViX-MangoEFormer: An enhanced vision transformer-EfficientFormer and stacking ensemble approach for mango leaf disease recognition with explainable artificial intelligence. Computers. 2025;14(5):171. doi:10.3390/computers14050171.

[44] Biswas R, Uddin J, Hasan MJ. A new approach of iris detection and recognition. International Journal of Electrical and Computer Engineering. 2017;7(5):2530-2536. doi:10.11591/ijece.v7i5.pp2530-2536.

[45] 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.

[46] Faruq O, Islam MI, Islam MS, Tarafder MTR, Rahman MM, Islam MS, Mohammad N. Re-imagining digital transformation in the United States: Harnessing artificial intelligence and business analytics to drive IT project excellence in the digital innovation landscape. Journal of Posthumanism. 2025;5(9):333-354. doi:10.63332/joph.v5i9.3326.

[47] 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.

[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;126:103667. doi:10.1016/j.inffus.2025.103667.

[49] Al-Mahmud O, Khan K, Roy R, Alamgir FM. Internet of Things (IoT) based smart health care medical box for elderly people. In: 2020 International Conference for Emerging Technology (INCET). IEEE; 2020. doi:10.1109/INCET49848.2020.9153994.

[50] Haque S, Chowdhury S, Faruq O, Akter R, Joy MSI, Munny MA, Shimu F. 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;12(1):915-923. doi:10.29284/v2jv8q59.

[51] Hossain MS, Khan A, Das P, Haque MSU, Kamruzzaman F, Akter S, Ahmed A, Miah MR. Enhanced market trend forecasting using machine learning models: A study with external factor integration. International Interdisciplinary Business Economics Advancement Journal. 2025;6(1):5-12. doi:10.55640/business/volume06issue01-02.

[52] 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.

[53] Siddiqui MIH, Rahman MS, 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;4(5):110-150. doi:10.32996/jmhs.2023.4.5.14.

[54] 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. Informatics in Medicine Unlocked. 2025;56:101657. doi:10.1016/j.imu.2025.101657.

[55] 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;18(3):294-307. doi:10.30574/wjaets.2026.18.3.0156.

[56] Rohan A, Tolie HF, Hasan MJ, Kannan S. Full autonomy in underwater robotics systems: A realistic prospect? Engineering Applications of Artificial Intelligence. 2025;162(Part C):112638. doi:10.1016/j.engappai.2025.112638.

[57] Adnan BM, Chakma S, Alam MMJ, Alamgir FM. Performance simulation and comparison in High Altitude Platforms communications systems under PSK, DPSK, QAM and FSK modulation schemes and AWGN, Rician and Rayleigh communication channels. In: 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON); 2016; Vancouver, BC. p. 1-11. doi:10.1109/IEMCON.2016.7746080.

[58] Naznin R, Sarkar MAI, Asaduzzaman M, Akter S, Mou SN, Miah MR, Sajal A. 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;5(11):21-32.

[59] Rahman MM, Gony MN, Ullah MS, Shuvra SMK, et al. MaizeFormerX: A lightweight vision transformer with cross-scale attention for explainable maize leaf disease diagnosis. Scientific Reports. 2026. doi:10.1038/s41598-026-44550-0.

[60] Arifeen M, Hasan MJ, Rohan A, Kannan S, Prathuru A, et al. Enhancing acoustic emission driven smart gas-pipeline monitoring with graph neural network. In: Manjurul Islam MM, Baptista ML, Tariq F, editors. Artificial Intelligence for Smart Manufacturing and Industry X.0. Cham: Springer; 2025. p. 165-178. doi:10.1007/978-3-031-80154-9_8.

[61] 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.

[62] Khan MDA, 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;6:931-942. doi:10.1109/OJCS.2025.3577588.

[63] Haque MM, Choudhury ZH, Tejesh GS, Alamgir FM. IoT based smart energy metering system for power consumers. In: 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET). IEEE; 2019. doi:10.1109/ICIET48527.2019.9290661.

[64] Shimu F. Intelligent cybersecurity framework: Machine learning-driven data protection and threat intelligence integration for modern digital communications. International Journal of Applied Mathematics. 2025;38(8s):620-632. doi:10.12732/ijam.v38i8s.595.

[65] Debnath J, Mohiuddin AB, et al. Hybrid vision transformer model for accurate prostate cancer classification in MRI images. In: 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE; 2025. doi:10.1109/ECCE64574.2025.11013952.

[66] Haque S, Islam H, Sharmin F, Joy MSI, Naher K, Rimi NN, et al. Generative artificial intelligence in enterprise information systems: Transforming business intelligence and strategic decision support processes. Journal of Information and Knowledge Management. 2025;20(2):887-897. doi:10.18848/8p0s2e25.

[67] Alamgir FM, Alam MS. A novel deep learning-based bidirectional Elman neural network for facial emotion recognition. International Journal of Pattern Recognition and Artificial Intelligence. 2022;36(10):2252016. doi:10.1142/S0218001422520164.

[68] 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;7(5):46-55. doi:10.32996/jmhs.2026.7.5.7.

[69] Tanim SH, Mithun MMU, Tarannum R. Sustaining vital care in disasters: AI-driven solar financing for rural clinics and health small businesses. American Journal of Technology Advancement. 2025;2(9):123-153. doi:10.31149/ajta.v2i9.2528.

[70] Karshiboev A, Al-Samad K, Tarafdar MTR, Rimi NN, Islam MS, Papel MSI. 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;12(1):387-410. doi:10.29284/9k2nx425.

[71] Majumdar J, Apu MH, Rahman M, Zaman T, Hassan MM. Multimodal EEG analysis of neural synchrony in minimal phrase processing using machine learning. Conference paper; 2025 Nov.

[72] Ahmed F, Alamgir FM. Simulation-based proportional study of routing protocols for MANET. International Journal of Computer Networks and Communications Security. 2017;5(12):28-36.

[73] Ghosh BP, Bhuiyan MS, Bishnu KK, Mahmud FU, et al. Personalized machine learning models for Parkinson's disease screening via voice biomarkers: Accounting for age, gender, and linguistic variability. The International Medicine. 2025 Dec.

[74] Debnath J, Uddin Khondakar Pranta AS, Hossain A, Sakib A, Rahman H, Haque R, Ahmed MR, Reza AW, Swapno SMMR, Appaji A. LMVT: A hybrid vision transformer with attention mechanisms for efficient and explainable lung cancer diagnosis. Informatics in Medicine Unlocked. 2025;57:101669. doi:10.1016/j.imu.2025.101669.

[75] Haque R, Sakib AA, Hossain MF, Islam F, Aziz FI, Ahmed MR, Kannan S, Rohan A, Hasan MJ. Advancing early leukemia diagnostics: A comprehensive study incorporating image processing and transfer learning. BioMedInformatics. 2024;4(2):966-991. doi:10.3390/biomedinformatics4020054.

[76] 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. 2025;17(20):9234. doi:10.3390/su17209234.

[77] 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.

[78] 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. 2026 Feb 20;29(2):114411. doi:10.1016/j.isci.2025.114411.

[79] Jashim FB, 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;15(02):1417-1431. doi:10.30574/ijsra.2025.15.2.1502.

[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] 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.