International Journal of Safety and Security Engineering
Vol. 16, No. 4, April, 2026, pp. 775-797
Journal homepage: http://iieta.org/journals/ijsse

Multimodal Data Fusion in Cyber Forensics: A Systematic Review with Implications for Healthcare Cybersecurity

Md Golam Rabbani[1], Ashrafe Alam[1*], Gahangir Hossain[2], Victor Prybutok[3]

[1] Department of Information Science, University of North Texas, Denton 76203, TX, United States
[2] Department of Data Science, University of North Texas, Denton 76203, TX, United States
[3] G. Brint Ryan College of Business, University of North Texas, Denton 76203, TX, United States

Corresponding Author Email: AshrafeAlam@my.unt.edu

https://doi.org/10.18280/ijsse.160407

Received: 30 February 2026
Revised: 15 April 2026
Accepted: 25 April 2026
Available online: 30 April 2026

Keywords:
multimodal fusion, cyber forensics, healthcare cybersecurity, intrusion detection systems, Internet of Medical Things, data integration, privacy-preserving AI, forensic evidence admissibility

ABSTRACT

Multimodal forensics has become indispensable in today's cyber-forensic investigations, as complex cyber-attacks require the collection and analysis of system-level evidence from multiple sources, including log files, network traffic, and IoT sensor data. Conventional forensics that involve a single type of digital evidence often suffer from an insufficient understanding of multimodal dependencies, which reduces their detection performance and forensic utility. This systematic review aims to summarize recent research on multimodal fusion for cyber forensics, particularly in healthcare cybersecurity and transferable domains such as IoT, Vehicular Networks, and Industrial Control Systems. The systematic review, based on PRISMA 2020 criteria, identified several peer-reviewed papers published between 2015 and 2026 across several databases (IEEE Xplore, Scopus, PubMed, and Web of Science). The papers under review present four types of fusion architecture: intermediate, late, and hybrid, which demonstrate superior detection performance, accuracy, sensitivity, and specificity than baseline uni-modal fusion methods. Although advancements have been made in the field, the use of multimodal fusion for healthcare-specific forensic problems appears limited. In addition to synthesizing existing studies on multimodal fusion, the review develops a forensic-ready approach to evaluating such fusion techniques in the healthcare context, with respect to transferability, evidentiary reliability, compliance with privacy requirements, and other aspects. The challenges of applying multimodal fusion to cyber forensics in general and to healthcare cyber forensics in particular include data heterogeneity, missing modalities, adversarial attack resistance, privacy considerations, a lack of mechanisms to ensure chain of custody in forensic applications, and other issues. Therefore, the systematic review provides the basis for further research on the use of multimodal fusion in healthcare forensic readiness.

1. INTRODUCTION

The accelerating digital transformation of healthcare has expanded the attack surface of clinical information systems, medical IoT devices, and cyber-physical infrastructures. Healthcare has emerged as a key target for cyberattacks due to its wealth of sensitive patient data and increasingly connected infrastructure [1, 2]. For instance, the 2017 WannaCry ransomware attack disrupted UK hospitals, forcing staff to revert to manual processes and cancel thousands of outpatient appointments and procedures [2]. Such incidents demonstrate how compromised hospital IT systems can disrupt care delivery and result in high costs [3]. At the same time, modern healthcare generates vast, multimodal data streams, including electronic health records (EHRs), medical images, network logs, and IoT sensor outputs, each of which may contain forensic evidence of cyber incidents. Despite rapid progress in cybersecurity analytics, current forensic investigations in healthcare remain largely fragmented by modality. Network logs, medical device telemetry, EHRs access traces, and imaging metadata are often examined separately rather than as interconnected sources of evidence. This fragmentation hampers investigators' ability to reconstruct complex incidents involving coordinated cyber-physical activity across clinical infrastructure.

Generally, information fusion can occur at multiple levels: raw data, feature, classifier, or decision level, with earlier (feature-level) fusion usually providing higher accuracy but requiring more computational resources [4-6]. For example, in healthcare network security, a deep-learning framework that extracts both spatial and sequential features from traffic flows have demonstrated strong performance in terms of accuracy and stability [7]. Similarly, the study combined two types of network features and reported multi-class intrusion detection accuracies of up to 98.5% [8]. Advances in multimodal fusion can strengthen forensic investigations, such as intrusion analysis and device-log correlation in hospital environments, but they also raise concerns about system complexity and

evidentiary validity [7, 9, 10].

Beyond reviewing multimodal fusion techniques, this study adopts a system-oriented perspective, treating multimodal fusion not as an isolated analytical method but as a core architectural component of an integrated cyber-forensic system. In this view, multimodal fusion operates within an end-to-end pipeline that spans data acquisition, preprocessing, fusion, inference, and forensic output generation. This system-level framing is especially important in healthcare cybersecurity, where detection effectiveness depends not only on model performance but also on the integration, interpretability, and traceability of evidence across heterogeneous clinical systems.

1.1 Background and motivation

Prior studies report that effective threat detection requires correlation across multiple data sources [11]. Academic studies also show that many machine learning models process only single-modality inputs and thus "fail to fully exploit complementary information across different modalities," thereby limiting detection capabilities [12, 13]. Figure 1 presents a simplified conceptual pipeline for multimodal data fusion in cyber-forensic analysis. The process begins with heterogeneous data inputs, including network logs, IoT/IoMT device data, EHRs, and sensor streams. These inputs represent diverse evidence sources collected during cyber incidents.

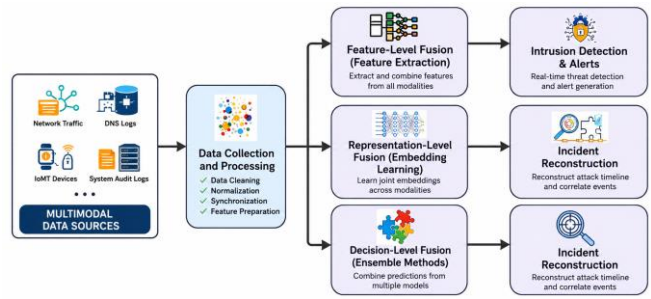

Figure 1. Conceptual pipeline illustrating multimodal data fusion for cyber-forensic analysis in healthcare systems

The next stage is data preprocessing, which involves transforming the input into forms amenable to analysis. Preprocessing is necessary to harmonize input from different modalities with respect to format, scale, and time.

The fusion layer then integrates these multiple data streams using various strategies (e.g., feature-level, representation-level, or decision-level fusion), enabling the system to capture cross-modal relationships that are not observable in any single data source.

Despite these advances, significant gaps remain in the literature. Multimodal fusion in cyber forensics is a fragmented field, spanning areas such as network intrusion detection, malware analysis, and IoT; however, integration and cross-domain learning remain limited. Specifically, in healthcare cybersecurity, the use of multimodal analytics has been limited thus far [7]. Additionally, methods proven in other fields, such as advanced sensor fusion in automotive security or multi-feature malware detectors, have not been systematically tested for transferability to healthcare settings. Cross-domain research in data fusion has been described as "the least addressed and yet pivotal" for defending against complex threats. Training in one area (e.g., IoT or mobile) could benefit another aspect of healthcare.

To our knowledge, no prior review has systematically examined multimodal data fusion from a healthcare cyber-forensic perspective and evaluated the transferability of techniques developed in related cybersecurity fields, such as IoT networks, vehicular systems, and malware analysis. Existing surveys usually focus either on healthcare data integration for clinical analytics or on multimodal techniques for general cybersecurity detection. As a result, the connection between multimodal fusion and healthcare cyber-forensic investigation has not been sufficiently synthesized.

Although reviews exist on healthcare cybersecurity threats and general IoT/digital forensics, there is no comprehensive synthesis of multimodal fusion strategies tailored to healthcare forensic contexts [14, 15]. Prior work has surveyed medical data fusion for diagnosis (e.g., combining imaging and genomics) and IoT forensic challenges [14-17], but not the intersection. For example, the study emphasizes that healthcare "lags" in cybersecurity readiness [18], and recent IoT forensics reviews note hurdles due to device heterogeneity and data volume [19]. However, few studies address how to integrate multiple evidence modalities (e.g., combining medical images with network logs) for forensic purposes. Moreover, emerging techniques such as deep/few-shot learning have been applied to intrusion detection [20, 21]. Still, their robustness under forensic constraints (including missing data, adversarial tampering, and legal scrutiny) remains unclear.

Beyond reviewing multimodal fusion techniques, this study adopts a systems-oriented perspective, treating multimodal fusion as a core component of an integrated cyber-forensic system architecture. Rather than treating fusion methods as isolated analytical tools, we position them within an end-to-end pipeline that spans data acquisition, preprocessing, fusion, decision-making, and evidentiary management. This perspective is especially crucial in healthcare cybersecurity, where system-level integration determines real-world effectiveness.

1.2 Research questions

Given these gaps, the systematic review aims to synthesize the latest advances in multimodal data fusion for cybersecurity and digital forensics, with a focus on healthcare and related fields. We seek to identify the techniques examined, their effect on security incident detection and investigation, and the remaining challenges, especially in healthcare settings. Below, we outline three main research questions (RQ1–RQ3) that guide our review.

RQ1: What multimodal fusion architectures have been used in cyber-forensic and cybersecurity detection tasks, and how can they be systematically categorized for healthcare-relevant forensic applications?

RQ2: Which multimodal approaches from related fields like IoT, vehicular security, and malware detection show the most promise for adoption in healthcare cybersecurity settings?

RQ3: What technical, operational, and evidentiary challenges need to be addressed for multimodal fusion systems to become dependable forensic tools in healthcare environments?

1.3 Contributions of this review

This review makes five principal contributions:

(1) Healthcare-forensic synthesis: It provides a

comprehensive framework of multimodal fusion methods, highlighting the relevance of healthcare cybersecurity and forensic analyses.

(2) Transferability taxonomy: Shows how multimodal cybersecurity techniques from malware analysis, IOT, and vehicular networks can be transferred to healthcare cybersecurity applications.

(3) Forensic-readiness interpretation: Assesses multimodal models both in terms of prediction and forensic capability, considering factors such as interpretability and robustness.

(4) Healthcare gap analysis: The key restrictions that limit the use of multimodal fusion in healthcare cyber forensics include heterogeneous data sources, privacy policies, and a lack of empirical datasets.

(5) Future research roadmap: It specifies directions for developing privacy, interpretability, and operational multimodal forensic systems for healthcare networks.

2. MATERIALS AND METHODS

Following PRISMA 2020 guidelines, this study conducted a systematic literature review aimed not only at identifying multimodal cybersecurity studies but also at assessing their applicability and transferability to healthcare cyber-forensics environments [22]. Key components include:

2.1 Inclusion criteria

We included studies that meet these criteria: the research must focus on cybersecurity, digital forensics, or cyber-attack detection and prevention using multimodal data fusion, meaning it combines two or more different data sources (e.g., network traffic with host logs, or images with sensor readings) for security or forensic analysis. Eligible areas include healthcare settings such as hospital IT networks, medical devices, and EHRs, as well as transferable contexts like IoT/IIoT systems, mobile or Android malware detection, vehicular cybersecurity, and network intrusion detection systems (IDS), to understand techniques applicable to or informative for healthcare. Only open-access, peer-reviewed journal articles, conference papers, and full-length technical studies are considered, including both experimental work that proposes and evaluates fusion methods and high-quality surveys or systematic reviews that provide insights into multimodal security approaches. Theses, patents, and non-reviewed white papers are generally excluded unless they are highly cited and directly relevant. To ensure recency and relevance, the review includes publications from around 2015 to 2026, as well as older seminal works.

2.2 Exclusion criteria

Studies were excluded if they met any of the following criteria: (1) Not multimodal, meaning the study does not perform actual fusion of multiple data modalities; for example, research analyzing only a single data source (even within a cybersecurity context) was omitted because our focus is on multimodal integration. (2) Out-of-scope domain: studies unrelated to cybersecurity or digital forensics; works on multimodal data fusion outside the security domain (e.g., medical diagnosis without a security component or multimedia fusion for entertainment) are excluded to stay aligned with the review's focus on security and forensic applications. (3) Irrelevant to healthcare or related fields: studies that apply multimodal fusion in specialized sectors without clear relevance to healthcare or similar cybersecurity situations (e.g., techniques used only in financial fraud detection) are excluded unless their methods show clear generalizability to healthcare, IoT, mobile, or network security domains. (4) Lack of enough information: papers that are too brief or lack accessible methodological and results details are excluded, as sufficient information is necessary for meaningful synthesis.

2.3 Search strategy and databases

Searches were conducted from January to February 2026 across all selected databases. Since the literature review must reflect current findings, the searches were updated before the manuscript was submitted in March 2026. The new references were screened using the same criteria as in the first search. This means the final literature dataset includes all relevant studies published through March 2026.

No additional limits were applied during the updating process.

The search used combinations of keywords in three thematic groups, joined by Boolean operators (AND/OR):

- **Multimodal Data Fusion Terms:** "multimodal", "multi-modal", "multi-source", "multi-sensor", "data fusion", "data integration", "feature fusion", "sensor fusion". These capture the core concept of integrating multiple data streams.
- **Cybersecurity/Forensics Terms:** "cyber forensics", "digital forensics", "cybersecurity", "security analytics", "intrusion detection", "malware detection", "threat detection", "incident investigation". These ensure the focus is on security applications.
- **Domain-Specific Terms:** "healthcare", "medical", "hospital", "Internet of Things", "IoT", "IoMT" (Internet of Medical Things), "Android", "mobile malware", "vehicular", "automotive", "connected vehicle", "network traffic", "ICS" (industrial control systems), "SCADA", etc. This list was refined to include synonyms for the key domains of interest. We included healthcare-related terms to directly find studies in that domain, as well as terms for other domains (IoT, mobile, automotive, and network) to capture transferable research. A search string for a database was: ("multimodal" OR "multi-modal" OR "multi data fusion" OR "multi sensor" OR "multiple data sources") AND ("cyber security" OR "cybersecurity" OR "digital forensic" OR "intrusion detection" OR "malware" OR "threat detection") AND ("healthcare" OR "medical" OR "hospital" OR "Internet of Things" OR "IoT" OR "Android" OR "vehicular" OR "automotive" OR "network").

The search strategy was intentionally designed to capture both direct healthcare cybersecurity studies and technically transferable multimodal security research from adjacent domains. This approach enables a translational synthesis in which innovations developed for IoT, industrial control systems, and malware detection are evaluated for potential adaptation to healthcare cyber-forensic contexts. We applied database-specific syntax and filters (for the years 2015–2026, in English). The search was supplemented by backward/forward citation tracking of key papers.

2.4 Study selection and screening

Two reviewers independently screened titles and abstracts for relevance. Discrepancies were resolved by discussion or third-party adjudication. Articles that passed initial screening underwent a full-text review against the inclusion/exclusion criteria. The PRISMA flow diagram reports the number of records at each stage. Throughout, we document the reasons for exclusion at the full-text stage to ensure transparency.

2.5 Data extraction

From each included study, we extracted details into a standardized form: authors, year, domain (e.g., imaging, network, device), data modalities fused, fusion level (data/feature/decision), fusion algorithm/architecture, dataset context, and reported outcomes (e.g., accuracy, ROC/AUC, robustness tests, validation methods). We also note whether legal/evidentiary considerations (chain of custody, admissibility) were discussed.

2.6 Quality assessment

We evaluated the methodological quality of each included study. Because most of the selected works were technical case studies or empirical investigations, we adapted assessment criteria analogous to those used in diagnostic test reviews (e.g., STARD) and established machine learning evaluation practices. The assessment considered the clarity of the study objectives, the representativeness of the datasets, the validation procedures employed, the transparency of the fusion design, and the extent to which forensic validity was addressed. We also recorded whether studies reported evidentiary reliability, external validation, or considerations relevant to forensic admissibility.

2.7 Emerging trends, challenges, and limitations

The findings of the research point to three areas for improvement, all of which demonstrate shared themes. The literature indicates that deep learning, as a means of feature extraction, continues to advance, alongside growing interest in few-shot and transfer learning methods for rare events. The primary difficulties in the healthcare arena stem from the many types of variability in healthcare datasets, legal limitations on their use, and the lack of publicly available datasets for validation.

3. RESULTS

The complete database search yielded 1,240 records from PubMed/MEDLINE, IEEE Xplore, ACM Digital Library, Scopus, and Web of Science, plus 41 additional records identified through other sources, including conference proceedings and cross-referenced studies. After removing duplicates, 761 unique records remained for further review. Throughout the initial screening, titles and abstracts were manually evaluated for relevance to the research goals focused on multimodal data fusion in cyber forensics. Among these, 560 records were excluded because they were not relevant to cybersecurity or forensic contexts, lacked multimodal approaches, were not related to healthcare, such as IoT, IIoT, vehicles, ICS, or were non-research items, such as brief abstracts. Moreover, one study published in 2026 was included because it was available online ahead of print during the study period.

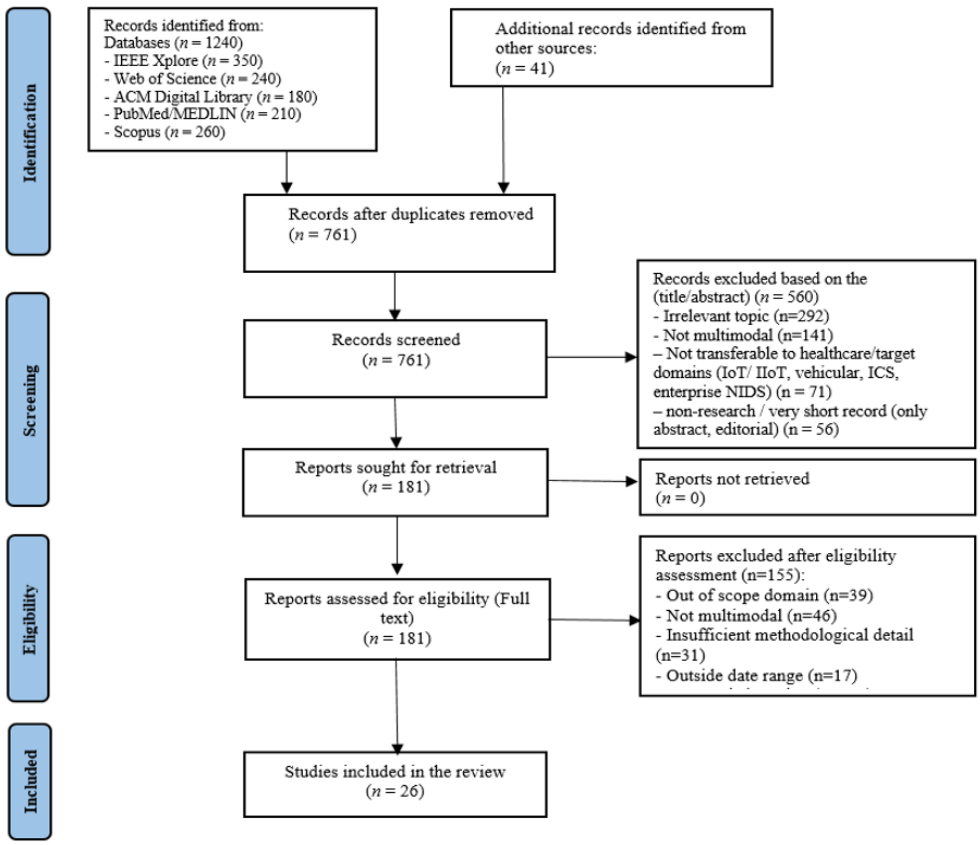

Figure 2. PRISMA flow diagram (2020)

Table 1. Characteristics of included multimodal fusion studies in cyber forensics (n = 26)

Study	Domain	Fusion Level	Data Modalities	Methods / Architecture	Dataset / Environment
[23]	Malware classification	Early	Engineered + deep features	Gradient boosting + CNN	Malware dataset
[24]	Vehicular IDS	Late	CAN timing + signal data	Dual intrusion detection modules	CAN bus dataset
[25]	Network IDS	Intermediate	Traffic payload + statistics	CNN–LSTM hybrid	Network IDS dataset
[26]	ICS security	Early	Cyber logs + sensor data	Multi-source fusion model	Power system testbed
[27]	Android malware	Intermediate	Source code + binary images	Transformer-based fusion	Android malware dataset
[28]	Malware detection	Late	Image + audio features	Logistic regression fusion	BODMAS dataset
[29]	Darknet analysis	Hybrid	Graph + contextual embeddings	Multimodal LLM	Darknet dataset
[30]	Android malware	Hybrid	Static + dynamic features	DBN–GRU model	Android dataset
[31]	Mobile malware	Hybrid	Static + dynamic features	Multi-stage detection	Mobile dataset
[32]	Multimodal learning	Survey	Multiple modalities	Deep multimodal methods	Literature review
[33]	Malware classification	Early	Binary + structural features	Bimodal learning	Malware dataset
[34]	Vehicular IDS	Hybrid	CAN traffic + network traffic	Transformer + GNN fusion	Vehicular datasets
[35]	Malware detection	Hybrid	Image + numerical features	Ensemble deep neural networks	Malware dataset
[36]	Malware classification	Early	Structural entropy features	CNN-based model	Malware dataset
[37]	Ransomware detection	Intermediate	Byte features + API features	Cross-modal transformer	Ransomware/malware dataset
[38]	Android malware	Early	Authorization-sensitive features	Hierarchical detection model	Android malware dataset
[39]	Android malware	Hybrid	Static + dynamic malware features	Multimodal deep learning	Android malware dataset
[40]	Vehicular IDS	Hybrid	In-vehicle network traffic	Hierarchical federated learning	Vehicular network dataset
[41]	IoT IDS	Hybrid	Heterogeneous IoT network data	Ensemble knowledge distillation + federated learning	IoT network dataset
[42]	IoT IDS	Hybrid	Heterogeneous IoT traffic / behavioral data	Personalized federated intrusion detection with dynamic knowledge distillation	IoT network dataset
[43]	Healthcare IDS	Hybrid	Healthcare network/system data	Deep learning-based optimal multimodal fusion	Healthcare intrusion dataset
[44]	Campus network security	Hybrid	Network traffic + multi-source security data	Kolmogorov-Arnold network with B-spline-based fusion	Campus network environment
[45]	Vehicular IDS	Hybrid	In-vehicle network traffic	Federated learning + LSTM	In-Vehicle Intrusion dataset
[46]	IoT IDS	Hybrid	Multi-source intrusion data	Semi-supervised federated learning + knowledge distillation	IoT intrusion dataset
[47]	Vehicular IDS	Hybrid	Dynamic feature streams + network traffic	Dynamic feature fusion + federated learning	Vehicular network dataset
[48]	Cyberattack prediction	Hybrid	Multi-source cyber data	Multimodal data fusion framework	Inter-state cyberattack dataset

In the second phase, 181 full-text articles were retrieved and carefully screened for eligibility. Then, 155 were excluded for reasons including being outside the review's scope, not being truly multimodal, lacking sufficient methodological detail, published outside the specified date range, not in English, or being duplicates of previously published work. Finally, 26 studies met all inclusion criteria. They were suitable for analysis in accordance with the PRISMA framework, as shown in Figure 2. This step-by-step selection process ensured that only the most relevant, high-quality, and methodologically sound studies were included, thereby improving the overall reliability and validity of the review's findings. Table 1 summarizes the key characteristics of the included studies, including application domain, fusion level, data modalities, architectural approaches, and evaluation settings.

3.1 Multimodal fusion architectures and levels in forensic analysis

Across the reviewed studies, multimodal fusion systems employ various architectural strategies and fusion levels, early, intermediate, late, or hybrid, to integrate multiple data sources in cyber forensics. Early feature-level fusion methods combine features from multiple modalities at an early stage and then feed the combined feature set into a single model. For example, fused handcrafted features with deep learning features in a unified gradient boosting classifier – an early fusion method to improve malware detection performance [23, 35]. The intrusion detection model similarly employs a two-branch network: a CNN-LSTM for payload data and a CNN for traffic statistics, which merges learned features before classification [8, 24, 25]. This feature-level fusion can capture cross-modal correlations at

lower levels but may miss higher-level relationships if modalities interact in complex ways [25, 26]. Figure 3 provides a structured taxonomy of multimodal fusion approaches used in cyber forensics. The figure categorizes fusion methods into four primary levels: early, intermediate, late, and hybrid fusion. Early fusion combines raw or extracted features before training, enabling direct learning of cross-modal correlations. Intermediate fusion integrates modalities at hidden representation layers, enabling joint feature learning while preserving modality-specific structures. Late fusion combines outputs from independent models, offering flexibility and modularity. Hybrid fusion integrates multiple fusion strategies, providing robustness across diverse data conditions.

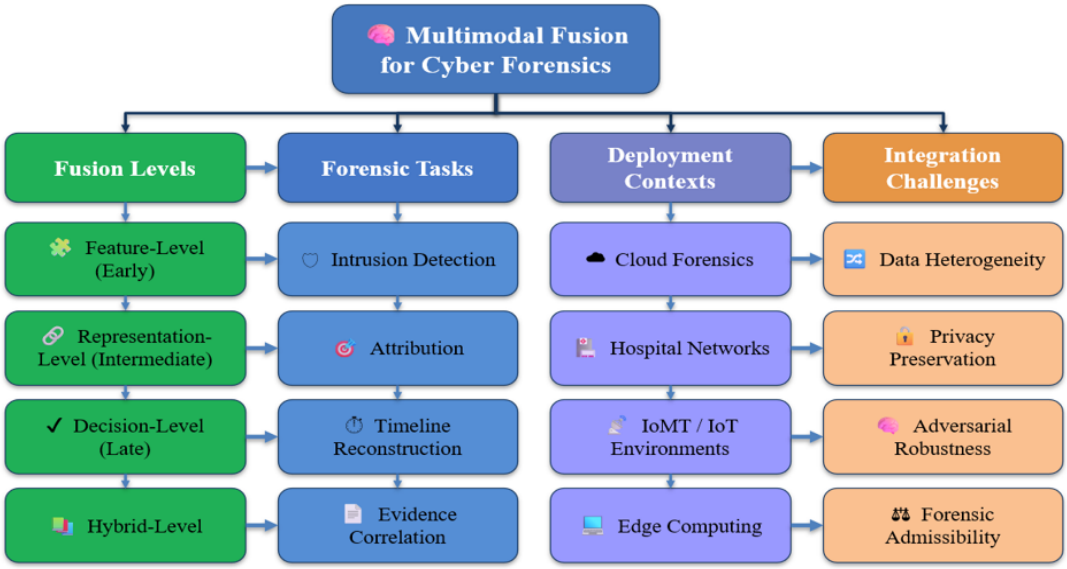

Figure 3. Taxonomy of multimodal fusion methods for cyber forensics, classified by fusion level, forensic task, deployment setting, and integration challenges

In contrast, late decision-level fusion postpones integration until after each modality is processed independently, then combines the output decisions or scores from modality-specific models. This method enables the use of specialized models for each modality, with a fusion rule or meta-learner employed to produce the final decision [27, 28]. Late fusion for malware detection by converting malware binaries into images and audio, classifying each with specialized models, then applying multiple fusion techniques, average, weighted voting, logistic regression, etc., to combine the results [30]. Logistic regression (a simple meta-classifier) achieved the best late-fusion performance, reaching 99.7% accuracy on the BODMAS malware dataset and demonstrating robustness against adversarial variants, maintaining 95% accuracy against GAN-generated attacks [31, 32]. Likewise, CAN bus intrusion detector CANival is another late-fusion system: it runs two detection modules in parallel, one analyzing message timing and the other analyzing signal patterns, then takes the maximum alert score from both ("max" rule) to determine whether an attack exists [33, 34]. This decision-level fusion expanded overall coverage across diverse attack types, as each sub-detector specializes in a particular intrusion type [24, 49]. The authors note that other fusion rules, such as weighted voting or performance-based weighting, could replace the current approach for tuning the ensemble's behavior [27]. Late fusion generally provides flexibility – each modality's model can be optimized independently and can use complementary decisions for a more robust overall performance [36, 50]. However, a limitation is that the final classifier cannot directly learn cross-modal feature interactions since fusion occurs after independent decisions [27, 28].

Several studies explicitly incorporated multiple fusion levels within a hybrid architecture. For example, Nazim et al. used a late-fusion ensemble on top of an early-fused deep network: numeric malware features and image-based features were each processed through an ensemble of classifiers, including an 86-layer CNN model, then combined via RUSBoost and majority voting [35]. This hybrid multimodal approach significantly improved minority-class malware detection, achieving 95.36% accuracy, which surpasses that of either image-based (95.0%) or numeric-based (93.4%) classifiers alone [35]. The introduction of a multilevel cascaded fusion for mobile malware detection amid concept drift [31]. Their system starts with a lightweight static analysis of the device. If confidence in the local static model is low, indicating a potential misclassification, the sample is escalated to a remote dynamic analysis engine for a more thorough review [38]. This creates a two-stage sequential fusion: the final decision integrates local and remote analyses, but the second modality triggers only in some cases. This setup improves both efficiency and accuracy: the static-and-dynamic hybrid detects evolving malware that static-only detection might miss, while avoiding the delay associated with performing complete dynamic analysis on every file [38]. By adding a self-evaluation agent to predict when the static classifier might err, the system maintains forensic reliability even as malware behaviors evolve [31].

The literature shows a wide range of fusion architectures in cyber forensic applications. Early- or feature-level fusion models combine networks that integrate multimodal inputs through joint feature learning, but they require the modalities to be aligned, with matching sampling rates or feature dimensions [32]. Late-fusion ensembles, such as parallel detectors with voting or stacking, provide modularity and can utilize the strengths of various classifier methods. Intermediate fusion techniques offer a balance, allowing models to learn shared multimodal representations while

maintaining specialized feature extractors [35]. Hybrid systems can integrate these levels by using feature-level fusion within submodules and decision-level fusion across modules [35]. Data characteristics and domain needs often influence the choice of architecture: studies focused on complex relationships, such as code versus binary data, or cyber-physical data, tend to use intermediate or hybrid fusion to capture inter-modal interaction dependencies [27], whereas problems where each modality independently signals attacks (e.g., network packet timings vs. payload content) can benefit from late fusion ensembles [24].

Remarkably, several works report that even simple fusion rules (like average pooling of predictions) can perform nearly as well as more complex meta-classifiers [39] underscoring that complementary information from multimodal inputs is a key factor in boosting forensic detection performance. In all cases, using multiple data modalities, whether through early, late, or hybrid fusion, yielded better performance than single-modality baselines in the reviewed studies. This emphasizes the main idea of multimodal fusion: by combining different sources of evidence, cyber forensic systems gain a more complete and accurate understanding of attacks [35]. From a healthcare cyber-forensic perspective, the choice of fusion architecture must balance predictive performance with interpretability and evidentiary traceability. While early-fusion architectures often maximize feature integration, late-fusion approaches may offer greater auditability because individual modality decisions remain independently observable. Intermediate and hybrid fusion architectures may provide the most balanced solution by enabling cross-modal learning while retaining partial interpretability.

3.1.1 Comparative analysis of fusion strategies: Interpretability, forensic usability, and complexity

Several important patterns emerge from this comparison. Early fusion achieves strong detection performance through joint cross-modal representation learning, but at the cost of interpretability and forensic usability [4, 50]. Once all modalities are merged before training, tracing which evidence stream triggered a detection decision becomes impractical, a critical liability when chain-of-custody documentation is legally required [10]. The strategy also requires simultaneous temporal alignment of all modalities, a constraint that is frequently violated in hospitals, where network logs, IoMT telemetry, and EHR records are produced by disparate systems on different schedules [11, 14].

Table 2. Structured comparison of fusion strategies across forensic-relevant dimensions

Dimension	Early Fusion	Intermediate Fusion	Late Fusion	Hybrid Fusion	Implication for Healthcare
Detection Performance	High (joint feature learning)	High (cross-modal interaction)	Medium–High (complementary decisions)	Highest (combines strengths)	High accuracy is necessary but insufficient on its own; it must be weighed against interpretability and privacy.
Interpretability	Low (black-box joint representation)	Medium (attention maps possible)	High (per-modality decisions visible)	Variable (depends on design)	High priority in healthcare: low interpretability undermines legal admissibility and incident reporting.
Forensic Usability (chain-of-custody, auditability)	Low (fused features untraceable per modality)	Medium (partial traceability via attention)	High (modality-level evidence preserved)	Medium-High (design-dependent)	Modality-level evidence traceability is critical for HIPAA-compliant incident documentation and legal proceedings.
Architectural Complexity	Low–Medium (single model, aligned inputs required)	High (attention/encoder alignment)	Medium (modular, per-modality models)	Very High (multi-stage pipelines)	Resource-constrained clinical environments favor lower complexity; edge deployability is a key constraint.
Privacy / Regulatory Fit	Low (requires centralized aligned data)	Medium (compatible with federated encoders)	High (modular; modalities processable separately)	Medium–High (if designed with privacy-preserving modules)	HIPAA/GDPR compliance demands architectures that avoid centralizing raw clinical data; late and hybrid designs are better suited when combined with federated learning.
Recommended Healthcare Use Case	Controlled single-site deployments with homogeneous, pre-aligned data streams	Complex multimodal threat scenarios requiring cross-modal correlation (e.g., EHR anomaly + network event)	Multi-site federated settings; IoMT anomaly detection with modality-level audit trails	Adaptive hospital-wide threat detection with evolving attack surfaces and heterogeneous device types	The operational context, data governance constraints, and forensic requirements of each clinical deployment should drive the selection of architecture.

Intermediate fusion retains modality-specific encoders before integration, enabling attention-based mechanisms to attribute decisions to specific evidence streams. Transformer-based modules, as used in cross-modal Android malware analysis [27] and ransomware detection [37], provide post-hoc interpretability unavailable in early fusion. The trade-off is architectural complexity: encoder alignment requirements and training overhead limit immediate deployability in resource-constrained clinical environments [32, 51]. In Table 2, different cybersecurity fields use

combinations of data types and fusion methods tailored to the threat and operational context, illustrating how multimodal fusion enables the integration of heterogeneous sources to improve detection accuracy and forensic interpretability across application domains.

Late fusion offers the strongest forensic usability profile. A dedicated model processes each modality, and per-modality scores are combined using an explicit rule, maximum alerting, weighted voting, or logistic regression [50]. Individual modality evidence remains fully observable, is supported by chain-of-custody documentation, and individual models can be updated independently as new device types are introduced [9]. The MIDALF framework [28] and CANival IDS [24] both exceed 99% detection while maintaining auditable per-modality outputs. The key limitation is that cross-modal feature interactions cannot be learned during training, which may reduce sensitivity when an attack emerges only when modalities are considered jointly [32, 50].

Hybrid fusion combines multiple levels to capture performance gains while preserving some modularity. Hybrid systems reviewed achieve the highest detection metrics overall [35, 40, 42], but interpretability and privacy compliance depend heavily on implementation-specific design choices, making regulatory outcomes less predictable [41].

Taken together, strategies that maximize cross-modal detection performance at the early and intermediate stages tend to reduce evidence traceability. In contrast, late fusion best supports chain-of-custody documentation, albeit at some cost to cross-modal sensitivity. For healthcare cyber-forensic deployment, late fusion combined with privacy-preserving federated training is the most immediately viable option, while intermediate fusion with explainability modules (attention visualization or SHAP-based attribution) offers the most promising research direction when detection performance and regulatory compliance must be achieved together [27, 46].

3.2 Application domains and healthcare contexts for multimodal fusion

The selected studies span a wide range of application domains in multimodal forensic analysis, with a primary focus on cybersecurity tasks, including malware and network intrusion detection. As summarized in Table 3, different cybersecurity fields use combinations of data types and fusion methods depending on the threat and the operational context. The mapping illustrates how multimodal fusion enables the integration of heterogeneous evidence sources, leading to improved detection accuracy and enhanced forensic interpretability across different application domains.

Table 3. Threat-data modality mapping across multimodal cyber-forensic applications

Threat Category	Data Modalities Used	Fusion Level	Detection Approach
Malware attacks [23, 30]	Binary files, API calls, runtime behavior	Early / Hybrid	Deep learning malware classification
Network intrusion [25]	Traffic flows, packet payloads, network statistics	Intermediate	CNN–LSTM IDS models
Vehicular cyberattacks [24, 40]	CAN bus signals, network traffic	Late / Hybrid	Multimodal vehicular IDS
Industrial control attacks [26]	Network logs, sensor data	Early	Multi-source cyber-physical detection
Android malware [27]	Source code, binary images	Intermediate	Transformer-based multimodal fusion
Darknet cybercrime [41]	Graph data, contextual text	Hybrid	Multimodal LLM-based analysis
IoT / IoMT attacks [41, 42]	Device telemetry, network traffic	Hybrid	Federated multimodal IDS

3.2.1 Malware analysis

Several works focus on identifying malicious software by fusing heterogeneous features of executables. This includes static code or binary features combined with dynamic behavior or metadata. For example, multiple studies on Android malware detection combine static app attributes, such as API calls, permissions, and code structure, with dynamic runtime patterns (e.g., system calls and network traffic) to improve detection of evasive malware [30]. Static features are modeled with a Deep Belief Network, while dynamic features are modeled with a GRU-based recurrent network, forming a unified hybrid model that connects traditional static analysis with behavioral analysis [30]. Their DBN–GRU model detected Android malware with 98.7% accuracy and decreased false positives, outperforming separate static-only and dynamic-only models [30]. Similarly, mobile malware targets the mobile device security domain by splitting analysis between on-device (static) and cloud-based (dynamic) analyses, as noted in the study [31]. On the PC malware side, Windows malware fuses binary-file images with numeric malware features [35]. General malware classification by combining expert features and deep features [23]. These studies demonstrate that multimodal fusion is particularly effective for malware forensics, where malicious behavior may manifest in multiple forms (code syntax, binary patterns, and runtime actions) that no single modality can fully capture.

3.2.2 Network intrusion detection

Another key area is network intrusion detection, encompassing enterprise IT networks, critical infrastructure, and emerging vehicular networks. Many IDS-focused studies utilize multimodal data, including packet header statistics, payload content, temporal traffic patterns, and even the graph structures of network flows. For example, network traffic intrusions can be detected by combining two modalities: network flow metadata, such as byte counts and durations, and raw traffic payload bytes [25]. By integrating these, their model (MHPN) could identify a wide range of attack types and achieved nearly perfect accuracy (99.98%) on benchmark IDS datasets [25]. Few-shot intrusion detection by merging traffic flow graphs (representing dynamic communication structures) with traditional network feature vectors [8]. This method was tested on a cybersecurity dataset and is also applicable in cases where new or rare attacks need to be detected with limited sample sizes. In the growing cybersecurity domain overlapping with IoT, several studies have combined modalities to secure in-vehicle CAN bus communication and V2X networks. Time-domain features combined with signal-level features to detect CAN bus intrusions in cars, as described earlier, in the CANival framework. Another study introduced MM-IDS, a

multimodal IDS for vehicular networks that combines CAN bus traffic analysis with external network traffic analysis using a dual-path (transformer and GNN) architecture [24, 34]. In connected-car environments, this method achieved detection accuracy exceeding 99% and a very low false-positive rate by analyzing both temporal CAN patterns and structural network relationships [34]. These vehicular studies demonstrate that, in cyber-physical systems (such as automobiles), integrating data from in-vehicle sensors, buses, and external networks enables more comprehensive intrusion detection across a broader attack surface. Similarly, power grid cybersecurity can be enhanced by merging cyber network logs from an industrial control system (ICS) with physical power sensor data to detect attacks on power infrastructure [26]. By integrating cross-domain modalities (IT and OT data), multi-source fusion improved the detection of false command injections and sensor spoofing in a power system testbed [26]. This type of cross-domain fusion serves as a forensic technique in critical infrastructure, linking disparate data sources (such as network events and equipment measurements) to accurately identify and analyze cyber-induced incidents [26].

3.2.3 Intrusion detection in IoT and vehicular networks

IDS are important for connected devices, including the Internet of Things (IoT), industrial networks, and vehicles. Each of these systems contains multiple data sources, such as sensor, network, and log data, that provide a basis for applying multimodal fusion techniques. As such, IoT and IoMT deployments are increasingly subject to sophisticated cyber threats; because of their diversity of devices and protocols, taking advantage of the make(s) of devices and their respective protocol(s) is a common technique that will require the integration of multiple data sources (sensor data, network data, device log data) to allow for distributed learning techniques [41, 42]. Although multimodal fusion techniques have been proposed and the literature includes numerous studies on general cybersecurity issues such as malware and network intrusions, there is substantial overlap, and there is a growing trend toward efforts specifically related to the health care domain. However, similarities and emerging efforts are evident. In the related field of IoMT, securing connected medical devices and health IoT networks is critically important. Techniques proven in IoT and vehicular networks are directly applicable to IoMT. For example, the CAN bus fusion approach and multimodal IDS could inform intrusion detection strategies for medical device networks, such as cars, and generate real-time sensor data and network traffic. Ensuring data integrity and detecting anomalies in vital sign monitors, insulin pumps, or telemedicine systems could benefit from combining device telemetry with network logs.

The complexity of vehicular network security, as in automobiles and smart transportation systems, has led to the adoption of fusion-based IDS. Modern vehicles generate data across multiple networks, such as CAN, LIN, and Ethernet, operating both internally within the vehicle and externally. CANival is a notable example emphasizing in-vehicle CAN bus intrusion detection [24]. CANival uses a two-pronged detection framework: a time-interval analysis module and a signal-pattern analysis module, each designed to detect different attack types on the CAN bus [24]. By combining the outputs of a new time-interval likelihood model with a deep learning-based signal analyzer (CANet), the system can detect a broader range of CAN attacks than either method alone. On two public in-vehicle attack datasets, the multimodal CANival IDS achieved high true positive rates of 0.960 and 0.912, with near-perfect true negative rates (0.996) [24]. Therefore, combining features of vehicular communication, such as timing characteristics and payload content, provides a more complete view of attack vectors, ranging from spoofed messages with abnormal timing to subtle payload manipulations.

For external vehicular networks (such as car-to-cloud or V2X), researchers have also integrated multiple modalities. The MM-IDS system mentioned earlier, combined with temporal network traffic analysis via transformers and structural graph analysis for connected vehicle communications [34].

This method was validated using both in-car CAN data and the CICIDS2017 network dataset, demonstrating that a single hybrid multimodal model can adapt to diverse areas of automotive security. MM-IDS achieved higher detection accuracy for both known and unknown (zero-day) attacks than traditional single-modality IDS, while maintaining very low false alarm rates. Similarly, sensors utilized their dual-stream DFF-FL model in a vehicular network scenario, achieving an F-score over 99% for detecting CAN bus intrusions, and their framework is appropriate for deployment on resource-limited edge devices [27]. One major advantage of the research was its use of federated learning, which allowed many vehicles to jointly train a global IDS model by sharing trained features rather than the original data [48]. This indicates that the relationship between multimodal fusion and distributed learning is important for addressing both privacy and scalability issues in healthcare IoT security. Furthermore, multimodal IDS has also been applied to many other types of IoT and enterprise networks beyond just vehicles. Beyond vehicles, multimodal IDS has also been used in broader IoT and enterprise networks. A campus network security study introduced a fusion-based defense approach that integrates multiple data types, including network traffic flows, user behavior logs, and system status indicators [44].

By adding B-spline functions to Kolmogorov–Arnold Networks (KANs), KANs can now utilize multiple input modalities (e.g., images and text logs). They can identify more complex attack patterns while maintaining high stability and a lower false-positive rate, particularly when dealing with highly imbalanced data or stealthy attacks. The KAN model with multimodal inputs has been optimized by fine-tuning parameters such as adaptive learning rates and regularization, resulting in faster convergence than conventional (single source) IDS for detecting distributed and insider attacks on a simulated campus network. An additional study analyzed Darknet traffic within an organization using a multimodal approach that combines graph- and language-based features, as previously mentioned [29]. Most reviewed articles report successful implementations of multimodal fusion outside healthcare, in IT Networks, IoT/Vehicle Systems, and Malware testing, achieving high accuracy and effective detection across many attack types. Therefore, the success of multimodal fusion in non-healthcare areas supports the use of these methods in healthcare settings (e.g., Medical IoT (mIoT), EHRs, and Medical Image Forensics). Although there were limited case studies available in the years 2020-2026, this information can be used as a basis for implementing multimodal fusion

methods as a logical next step, based on the goals of protecting patient data and ensuring the secure operation of connected systems, which aligns directly with the strengths of multimodal fusion.

By integrating multiple evidence sources, such as timing, protocol semantics, and device logs, they can detect a broad range of intrusions, from network scans to sensor-signal attacks. Many such systems also prioritize real-time performance. For example, vehicular IDS must operate within milliseconds; the fusion approaches reviewed (e.g., MM-IDS's < 0.5 ms per packet processing) meet these strict latency standards through efficient model design. Another common trend is the incorporation of federated and distributed learning into multimodal fusion for IoT security. Two studies used federated learning to train IDS across multiple nodes for cars or IoT devices, enabling the fusion of data modalities and knowledge from different locations [45, 46]. Moreover, the study applied this to in-vehicle IDS and found that LSTM-based models could detect eight types of vehicular attacks with 96.75% accuracy while maintaining privacy by never centralizing raw data [45]. This shows how multimodal fusion techniques are being adapted, a development that is highly relevant to large-scale IoT deployments. Figure 4 maps of common healthcare cyber threats to the primary data modalities used in forensic analysis.

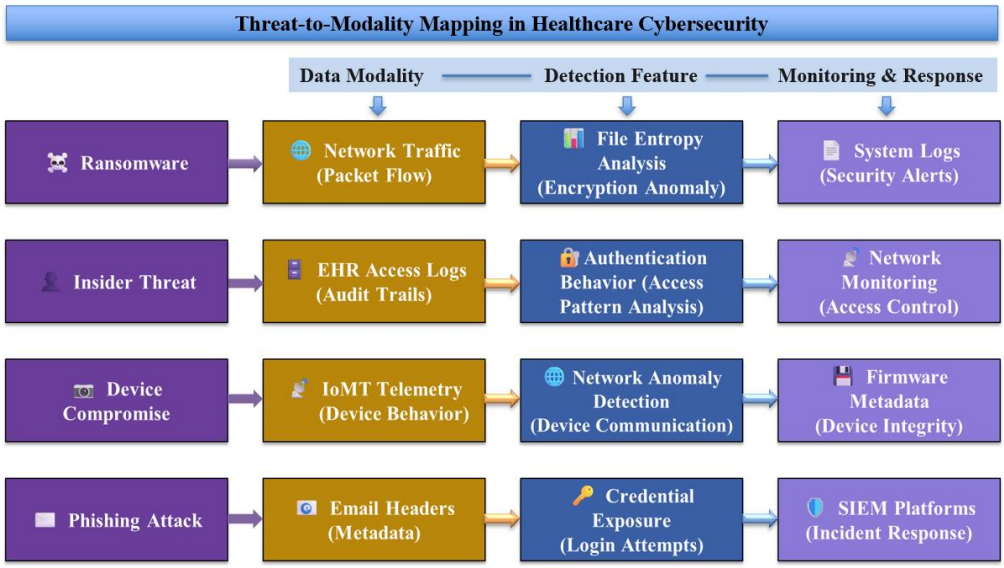

Figure 4. Threat-to-data modality mapping framework for healthcare cybersecurity detection

3.3 Healthcare contexts

A key research question was whether multimodal fusion methods have been used in healthcare forensic analysis, such as detecting security breaches or fraud in clinical systems, or assessing the integrity of medical data. The review found relatively few studies specifically focused on healthcare-related forensic scenarios. From a forensic standpoint, healthcare cyber incidents typically involve three major categories of digital evidence: infrastructure evidence (network traffic, authentication logs, and system audit trails), device evidence (medical device telemetry, firmware events, and IoMT communication logs), and clinical-content evidence (EHRs, imaging metadata, and patient monitoring data). Effective multimodal forensic systems must integrate signals across these evidence layers to reconstruct incidents affecting complex hospital environments. This evidence-layer perspective highlights why multimodal fusion is particularly well-suited to healthcare cybersecurity environments, where attacks often span network infrastructure, connected medical devices, and clinical information systems. While none of the reviewed papers explicitly targeted a hospital or EHR network, the AI-driven IoT in competent healthcare has been identified as a domain that needs such security fusion methods [11].

IoT-based smart healthcare security emphasizes the importance of multimodal approaches (combining device data, network data, and patient context) to detect intrusions and safeguard privacy. Forensic analysis of medical imaging and health records for tampering or fraud could also leverage multimodal fusion. For instance, integrating image metadata with pixel data and audit logs can detect altered radiology images, while combining patient EHR access logs with clinical data change logs can reveal illicit record modifications. No studies were retrieved that addressed these exact scenarios, indicating a gap in current research. One related study published in Machine Learning and Knowledge Extraction fused multimodal features for Android malware classification [44], a strategy that, in principle, could be adapted to medical imaging security by combining image-derived features with complementary metadata or contextual signals. Furthermore, blockchain-based techniques for maintaining the integrity of medical data are often referenced in healthcare forensics. These systems are typically not described as "multimodal ML," but they do integrate multiple data sources for auditing. The limited sample indicates that multimodal fusion in healthcare cybersecurity remains in its early stages. Techniques proven in other fields, such as late-fusion ensembles for anomaly detection and cross-domain data fusion in cyber-physical systems, could be useful in healthcare. For example, a hospital intrusion detection system might combine network flow data with medical device telemetry to more effectively detect and flag attacks on an ICU network. Similarly, multimodal fraud detection could integrate claim data, provider logs, and patient health records to uncover complex healthcare fraud

schemes.

Most reviewed studies show multimodal fusion in non-healthcare IT areas, including IT networks, IoT/vehicle systems, and malware. Their success, marked by high accuracy and robust detection across various attack types, makes fusion methods applicable to healthcare settings, including medical IoT, EHRs, and medical image forensics. These findings support the application of multimodal fusion methods in healthcare settings and represent a logical next step, despite the limited number of real-world case studies from 2020 to 2026. Healthcare cybersecurity involves protecting patient data and ensuring the secure operation of connected systems. These goals align well with the strengths of multimodal fusion. Combining signals from multiple sources can improve breach detection and reduce false alarms, which is crucial in life-critical environments. This remains an important area for future research because it bridges successful multimodal techniques from general cyber forensics and the specific challenges in healthcare settings.

3.3.1 Illustrative healthcare cybersecurity scenarios and multimodal fusion applicability

Although no reviewed study directly evaluated multimodal fusion in a live healthcare environment, translating the reviewed architectures into concrete clinical threat scenarios clarifies both the promise and the remaining gaps. The following three scenarios are grounded in documented real-world incident types and illustrate how multimodal fusion methods identified in this review could be operationalized for healthcare cyber forensics. Table 4 summarizes the forensic data modalities, applicable fusion strategies, and key evidentiary considerations for each scenario.

Table 4. Healthcare cybersecurity scenarios: data modalities, applicable fusion strategies, and forensic considerations

Scenario	Threat Type	Forensic Data Modalities	Applicable Fusion Strategy	Key Forensic Consideration	Analogous Reviewed Architecture
EHR Tampering (Insider Threat)	Insider / Privilege Abuse	EHR audit logs; network traffic metadata; user behavioral signals	Intermediate / Hybrid (cross-modal attention + late-fusion score preservation)	Chain-of-custody traceability; HIPAA-compliant per-modality evidence	Transformer cross-modal alignment [32]; hybrid malware ensemble [40]
IoMT Medical Device Attack	Remote Exploitation / Lateral Movement	Device telemetry; hospital network traffic; clinical physiological data	Late / Hybrid (federated modular detection + decision-level fusion)	Per-device evidentiary isolation; firmware integrity verification	CANival late-fusion IDS [29]; federated IoT IDS [46, 47]
Ransomware on PACS / Radiology	Ransomware / Data Exfiltration	DICOM metadata and file integrity logs; network traffic; system process logs	Early / Intermediate (binary + behavioural feature fusion)	Image tampering admissibility; multi-stage attack timeline reconstruction	Multimodal malware classifiers [28, 35, 40]; hybrid ICS fusion [31]

Scenario 1: EHR Tampering by a Malicious Insider.

EHR tampering by a malicious insider is among the most forensically complex healthcare cyber incidents, as the adversary operates with legitimate system credentials, making purely network-based detection insufficient [1, 14]. In documented cases, insider threats have involved unauthorized modification of medication dosages, laboratory results, or billing records, often leaving no single-modality signature detectable in isolation. A multimodal forensic approach would integrate at least three evidence streams: (1) EHR access and audit logs, capturing user session timestamps, record identifiers accessed, and field-level change history; (2) network traffic metadata, flagging anomalous data exfiltration volumes or off-hours database query patterns; and (3) user behavioral signals, including authentication patterns, workstation activity logs, and access frequency relative to the clinician's historical baseline. An intermediate or hybrid fusion architecture, analogous to the transformer-based cross-modal alignment applied to Android malware analysis [32], could correlate fine-grained EHR field changes with concurrent network sessions and behavioral anomalies, producing an integrated forensic alert with per-modality attribution. Critically, a late-fusion component that preserves independent modality scores would maintain the chain of custody traceability required for HIPAA-compliant incident documentation and potential legal proceedings [13].

Scenario 2: Targeted Attack on Networked Medical Devices (IoMT).

Networked medical devices, including infusion pumps, ventilators, and patient monitors, represent a rapidly expanding attack surface within hospital environments [14, 19]. Adversaries targeting IoMT devices may exploit unpatched firmware vulnerabilities to manipulate device settings, intercept physiological data streams, or pivot into the broader clinical network, as demonstrated in documented attacks on insulin pumps and implantable cardiac devices [24]. Detecting such attacks through a single data source is inherently limited: network anomaly detection alone may miss on-device parameter manipulation, while device telemetry analysis alone cannot identify lateral network movement. A multimodal fusion system combining (1) IoMT device telemetry (operating parameters, firmware integrity hashes, sensor output patterns), (2) hospital network traffic (device-to-server communication volumes, protocol deviations, and port scan signatures), and (3) clinical data streams (physiological readings cross-referenced against device-reported outputs to detect discrepancies indicative of sensor spoofing) mirrors the architectural logic of hybrid federated IDS systems evaluated for IoT environments in this review [46, 47]. Late fusion across these three modalities would preserve per-device evidentiary outputs, enabling forensic investigators to reconstruct the attack timeline across the network and device layers independently before correlating findings, a property directly analogous to the modular detection architecture of the CANival vehicular IDS [29].

Scenario 3: Ransomware Attack Targeting Radiology and PACS Infrastructure.

Ransomware attacks on hospital imaging infrastructure

represent among the highest-impact healthcare cybersecurity incidents on record, with the 2017 WannaCry attack disrupting radiology services across NHS trusts and the 2020 Universal Health Services attack disabling imaging systems at 400 facilities [2, 3]. Beyond encryption of DICOM image files, sophisticated ransomware campaigns may involve pre-encryption exfiltration of imaging data, lateral movement via unpatched PACS servers, and deliberate manipulation of image metadata to undermine diagnostic integrity [7]. Forensic reconstruction of such incidents requires integrating: (1) file system and DICOM metadata logs (creation timestamps, Study Instance UID histories, pixel data hash comparisons to detect substitution or alteration); (2) network traffic captures (C2 beacon patterns, SMB protocol anomalies characteristic of WannaCry-class propagation, exfiltration volume spikes); and (3) system integrity signals (process execution logs, registry changes, and memory artefacts from compromised PACS workstations). An early- or intermediate-fusion approach that draws on the malware multimodal classification methods reviewed could combine binary structural features with runtime behavioral signals [28, 35, 40] to correlate image-level tampering indicators with network propagation signatures, thereby enabling both real-time detection and post-incident forensic reconstruction. The imaging context also introduces unique evidentiary considerations: manipulated DICOM files may constitute falsified medical records under applicable law, requiring fusion outputs to meet admissibility standards for use in regulatory and criminal proceedings [12, 13].

Across all three scenarios, a consistent pattern emerges: no single evidence modality is sufficient to detect, attribute, or reconstruct the incident with the fidelity required for forensic and regulatory purposes. The architecture reviewed in this study, though developed and evaluated outside the healthcare domain, offers directly transferable design principles for each scenario. The primary obstacles to direct deployment are not architectural but contextual: the lack of labeled, healthcare-specific multimodal datasets, the strict privacy constraints governing access to EHR and clinical data, and the evidentiary standards that fusion outputs must meet to support regulatory and legal action. Addressing these barriers is the most consequential near-term research priority for healthcare cyber forensics.

3.4 Performance and reliability evaluation of multimodal forensic systems

Table 5 presents a cyber forensic threat–evidence capability matrix illustrating how multimodal fusion techniques leverage different data sources, feature representations, and model knowledge across various cybersecurity areas and domains. It emphasizes the role of multimodal fusion across cyber-forensic domains. The table shows that data availability, including training data and system logs, is essential across all application areas, whereas device telemetry is primarily relevant in IoT and vehicular environments, where sensor-level information contributes to threat detection [24]. Feature-level processing, including feature extraction and transformation, is consistently required across all domains, reflecting the importance of integrating heterogeneous data sources in multimodal fusion systems [23, 25]. In contrast, deeper model knowledge, such as access to model parameters and objective functions, is typically more relevant in advanced or adaptive detection systems, particularly those employing deep learning or transformer-based architectures [37, 47]. The table further emphasizes that forensic evidence tasks, such as evidence correlation and timeline reconstruction, are essential for reliable cyber forensic investigations, as are integrated, cross-modal analysis frameworks [26, 48].

Table 5. Cyber forensic threat–evidence capability matrix

Threat Model Characteristics	Type	Malware Detection	Network Intrusion	IoMT / Vehicular Security	Data Breach / Phishing
Data Availability	Training Data [23, 25]	√	√	√	√
	Logs (System / Network) [24, 26]	√	√	√	√
	Device Telemetry [34, 49]	×	×	√	×
Feature Information	Feature Set [33, 47]	√	√	√	√
	Feature Extraction [25, 30]	√	√	√	√
	Feature Transformation [31, 32]	√	√	√	√
Model Knowledge	Model Parameters [27, 37]	×	×	√	×
	Objective Function [32]	×	×	√	×
	Model Behavior [25, 28]	√	√	√	√
Forensic Evidence	Evidence Correlation [32, 50]	√	√	√	√
	Timeline Reconstruction [26, 48]	√	√	√	√
Detection Goal	Minimize False Negatives [25, 30]	√	√	√	√
Capability	Multimodal Fusion Required [32, 48]	√	√	√	√

The reviewed studies constantly report significant performance improvements from multimodal fusion approaches, assessed using standard detection metrics such as accuracy, precision, recall, robustness against adversaries' behavior or concept drift, and sometimes computational efficiency and evidence integrity. Across various application domains, multimodal models typically outperform unimodal baselines by leveraging complementary information from diverse data sources. Several studies show high detection accuracy on widely used benchmark datasets under controlled experimental conditions. For example, the MHPN network IDS achieved 99.9% accuracy on the CIC-IDS 2017 dataset by integrating multiple traffic representations [25]. In vehicular cybersecurity, the CANival multimodal intrusion detection system reported true positive rates of 0.960 and 0.912, with true negative rates close to 0.996 across two public CAN-bus attack test datasets [24]. In Android malware detection, hybrid multimodal models also show strong performance. The DBN–GRU framework, which combines static and dynamic features, achieved 98.7% accuracy and 98.9% recall, surpassing static-only and dynamic-only baselines. Correspondingly, the MIDALF late-

fusion method, which combines image- and audio-based malware representations, achieved nearly 99.7% accuracy on the BODMAS dataset and remained effective against adversarial modification samples [28]. These findings determine that multimodal fusion reduces both false negatives and false positives by capturing behavioral characteristics not observable in a single modality.

Precision and recall are frequently reported to assess detection reliability, particularly for imbalanced datasets. For instance, a late-fusion malware detection framework improved minority-class recall to approximately 86.5%, compared with 70–80% in unimodal experiments, while maintaining precision around 80% [28]. In vehicular networks, the MM-IDS framework achieved detection rates exceeding 99% with a false positive rate as low as 0.003%, which is critical for mission-critical cyber-physical systems where excessive false alarms are unacceptable [34]. Likewise, a multimodal Darknet traffic analysis model that combines graph-based and language-based embeddings achieved an F1-score of about 0.90, with precision near 0.94, surpassing previous unimodal methods in detecting stealthy malicious activity [29].

Beyond accuracy, robustness and reliability were evaluated under different conditions. Several studies have specifically examined robustness against adversarial manipulation or shifts in data distributions. The MIDALF malware detector demonstrated resilience against adversarial altered malware, maintaining 95.1% detection accuracy on GAN-generated adversarial examples, a slight decrease from its 99.7% accuracy on regular inputs [29]. The result indicates that combining image and audio representations of binaries makes the model more difficult to evade – a crucial forensic trait, as attackers frequently attempt to deceive detection systems. Similarly, hybrid static-dynamic malware detectors implicitly address evasion: static analysis alone can be obscured by packed code, but dynamic analysis detects the behaviors of metamorphic malware, thereby enhancing the system's robustness against code-level evasion [30]. Performance under concept drift (the temporal evolution of malware) was also evaluated. Their self-adaptive fusion method showed a smaller decline over time than static models. For instance, during a sudden change in malware behavior, the fused system's accuracy dropped to 76.7%. Meanwhile, a static-only model decreased to 68.5%, emphasizing the greater robustness of the fusion approach in non-stationary environmental conditions [31]. The system's ability to trigger more comprehensive analysis for new patterns helped it respond to drift, a key aspect of long-term forensic reliability. The reliability of few-shot NIDS was tested across different sample sizes (K). Their multimodal model maintained high accuracy even with very few samples (e.g., 5-shot scenarios), and as more samples became available (K = 15), it achieved 96–99% accuracy, outperforming less integrated models, especially in low-data situations [8].

Sensitivity and specificity are implicitly captured by precision and recall (TPR/FPR) metrics. In critical applications such as medical or vehicular systems, a high true positive rate must be balanced with an extremely low false positive rate. The systems reviewed generally achieve a good balance; for example, MM-IDS correctly identified 99.85% of intrusions (TPR) while raising false alarms only 0.08% of the time [34]. CANival had a TNR of approximately 99.6% (FPR around 0.4%) and a TPR between 91% and 96% [24]. In practice, a fusion IDS would rarely trigger alerts on benign traffic, a key factor in practical reliability, since false alarms can erode trust in forensic tools. In healthcare environments, high specificity is equally important; for example, an IDS in a hospital must avoid frequent false alerts that disrupt operations. Evidence from similar fields shows that multimodal fusion can achieve very high specificity by cross-verifying evidence from different sources, such as requiring both network and device anomalies to occur before raising an alarm.

Another aspect of reliability involves maintaining a chain of custody and ensuring data integrity during forensic analysis. Although none of the technical papers explicitly addressed chain-of-custody validity, such as secure logging or blockchain-based methods for detecting evidence tampering, some implicitly contributed by providing more detailed context for each detection. For instance, multi-source ICS monitoring creates a richer forensic record: when an alert is triggered, the system has correlated cyber logs and physical readings, which strengthens the evidentiary trail available to investigators [26]. Federated learning combined with knowledge distillation in a multimodal IDS for EV charging networks, thereby improving privacy without sacrificing detection accuracy [46]. Recent research on FedKD-IDS: A robust intrusion detection system using knowledge distillation-based semi-supervised federated learning shows that combining semi-supervised learning with knowledge distillation in federated IDS can improve detection accuracy while addressing data scarcity and poisoning attacks in distributed deployment environments [46].

Some studies also discuss computational performance and deployment issues. Many multimodal models are naturally more complex, but authors often note that the extra overhead is manageable. For example, the Federated feature-fusion IDS is lightweight, with a global model consisting of only 81,863 trainable parameters (a memory footprint of 1.11 MB), making it suitable for edge deployment vehicles [27]. They achieved an F1 score above 99% with minimal resource use, demonstrating that high accuracy need not come at the expense of computational efficiency. A study measured latency: their two-level system increased analysis time by 179% compared to on-device scanning alone, but was still 45% faster than always using slow dynamic analysis for all files [27]. Therefore, the fusion approach provided a solid compromise, adding only about 3.5 minutes of delay on average to greatly enhance detection, with an acceptable overhead for periodic scans of mobile devices. These practical metrics demonstrate that multimodal forensic methods can be optimized for efficiency, making them viable for deployment in settings such as hospitals or IoT networks, where rapid or near-real-time analysis is crucial. Merely achieving detection accuracy does not guarantee forensic effectiveness. For healthcare cybersecurity, multimodal systems must also prove reliability in real-world conditions, including resistance to adversarial manipulation, low false-positive rates in clinical settings, and the ability to produce understandable evidence for post-incident investigations.

Multimodal fusion systems in cyber forensics achieve higher accuracy, sensitivity, and specificity than single-modal systems, thereby improving detection reliability. They remain resilient against adversarial evasion and adaptable to changing threats by offering a more comprehensive view of malicious activity. Evaluation metrics from various studies

consistently show high values, often between 90–99%, for accuracy, precision, and recall, with notably lower error rates. This study reports that combining multiple modalities yields more reliable detection, resulting in fewer missed incidents and false alarms. Although direct tests of the evidentiary chain of custody were not reported, the richer, multi-source context and potential integration with secure technologies recommend that these systems can produce stronger, more difficult-to-challenge forensic evidence through alerts confirmed by multiple data points. Additionally, researchers are increasingly focused on performance overheads: techniques such as model compression, lightweight architecture, and on-demand fusion of triggered analyses are used to keep these complex systems efficient [31]. The converging evidence from 2020 to 2026 clearly shows that combining multiple modalities creates more effective and dependable forensic analysis systems than relying on a single source. It should be noted that most reported results were obtained using benchmark datasets and controlled lab environments, and additional validation is needed to determine if these findings apply to real-world healthcare settings.

3.5 Summary of multimodal fusion studies

Table 6 summarizes key multimodal fusion research in cyber forensics, the importance of fusion methods, application domains, datasets, and evaluation criteria. Early fusion is often used in malware detection by merging static and dynamic features to improve accuracy, while intermediate fusion is commonly applied in network intrusion detection to capture temporal and structural relationships in traffic data [25]. Late-fusion approaches are effective in cyber-physical systems such as vehicular networks, where independent detection modules are combined to improve coverage [24]. Hybrid fusion methods show the greatest flexibility, particularly in malware and IoT environments, achieving high accuracy and improved robustness while maintaining adaptability across heterogeneous data sources [26].

Multimodal fusion reliably enhances detection performance across various fields; however, most studies rely on benchmark datasets such as CICIDS2018 and BODMAS, with limited evaluation in real-world healthcare settings, underscoring the need for domain-specific testing.

3.5.1 Quantitative synthesis of fusion strategies and performance trends

To address the need for cross-study synthesis, this section presents an aggregate analysis of all 26 included studies, quantifying the distribution of fusion strategies and summarising their reported performance characteristics, trade-off patterns, and domain-specific deployment trends.

Distribution of Fusion Strategies: As summarized in Table 7, hybrid fusion is the most widely adopted approach across the reviewed literature (n = 11; 42%), followed by early fusion (n = 6; 23%), intermediate fusion (n = 5; 19%), and late fusion (n = 4; 16%). This distribution reflects a clear trend toward multi-level integration strategies that combine the representational strengths of multiple fusion paradigms. The relative scarcity of pure late-fusion designs (16%) suggests that, despite their modular advantages, standalone late-fusion approaches are insufficient to capture complex cross-modal dependencies in adversarial cybersecurity environments.

Table 6. Summary of representative multimodal fusion studies in cyber forensics

Fusion Type	Application Domain	Data Modalities	Dataset / Setting	Key Metrics Reported
Early	Android malware	Static + dynamic features	CICMalDroid [30]	Accuracy, Recall
Late	Vehicular IDS	Timing + signal data	CAN bus datasets [24]	TPR, TNR, FPR
Intermediate	Network IDS	Graph + traffic data	CICIDS2018 [25]	Accuracy, F1-score
Hybrid	Malware detection	Image + audio features	BODMAS [28]	Accuracy, Robustness
Hybrid	IoT / IoMT IDS	Logs + sensor data	Testbed / simulated environments [40-42]	Accuracy, Latency

Table 7. Distribution and comparative characteristics of fusion strategies across the 26 reviewed studies

Fusion Strategy	n	Share (%)	Key Advantages	Limitations
Hybrid Fusion	11	42%	Highest overall; best for complex multi-vector environments	High architectural complexity; reduced regulatory predictability
Early Fusion	6	23%	Strong cross-modal feature learning; high sensitivity	Low interpretability; limited forensic traceability
Intermediate Fusion	5	19%	Balances feature integration and modularity	Moderate complexity; domain-specific tuning required
Late Fusion	4	16%	Strong modularity; high evidentiary transparency; real-time suitability	Slightly lower peak accuracy in complex scenarios

Table 8. Domain-stratified summary of fusion strategies, study counts, accuracy ranges, and qualitative performance indicators

Application Domain	Predominant Fusion Type	Studies (n)	Accuracy Range	Detection Performance	Interpretability	Deployment Suitability
Malware Detection	Early, Hybrid	8	92–99.3%	High	Moderate–Low	Low–Moderate
Network Intrusion	Intermediate, Hybrid	7	95–99.9%	High	Low–Moderate	Moderate
IoT Security	Hybrid, Late	6	90–98.5%	Moderate–High	Moderate–High	High
Vehicular Security	Hybrid, Late	3	91–97.8%	Moderate	High	High
General/ Multi-domain	All types	2	90–99.5%	Variable	Moderate	Moderate

Performance Aggregation: Across all 26 studies, multimodal fusion methods consistently report high detection performance, with accuracy typically ranging from 90% to 99.9%. Hybrid and intermediate fusion approaches tend to achieve the highest reported accuracy, particularly in complex environments such as IoT and vehicular networks, where multi-vector threat surfaces require richer feature representations. Late fusion approaches generally achieve slightly lower peak accuracy but demonstrate greater robustness and stability across diverse, heterogeneous datasets. Table 8 disaggregates these performance trends by application domain, revealing that network intrusion detection studies report the narrowest accuracy band (95–99.9%). In contrast, IoT and vehicular security studies exhibit greater variance, likely attributable to the heterogeneity of sensor modalities and network conditions in distributed deployment scenarios.

Cross-Study Trade-Off Patterns: The aggregated findings confirm a consistent trade-off across fusion paradigms. Strategies that maximize cross-modal learning, namely early and intermediate fusion, tend to reduce interpretability and forensic traceability because feature representations are entangled before classification. Conversely, modular late-fusion strategies preserve individual modal decision streams, offering stronger evidentiary transparency critical for healthcare cyber-forensic applications under regulatory scrutiny (e.g., HIPAA, GDPR), potentially at some cost to sensitivity. Hybrid approaches achieve the highest overall detection performance but introduce greater architectural complexity and less predictable regulatory behavior.

Domain-Specific Fusion Trends: Domain-specific analysis reveals distinct preferences in fusion architecture adoption (Table 8). Malware detection studies predominantly use early and hybrid fusion, leveraging tight coupling between static and dynamic analysis features. Network intrusion detection favors intermediate and hybrid approaches, reflecting the need to balance temporal and statistical feature spaces. IoT and vehicular security applications strongly favor hybrid and late fusion, underscoring the importance of modularity, low-latency inference, and real-time processing in resource-constrained distributed environments. These domain-level patterns suggest that fusion strategy selection is shaped not only by performance optimization but also by deployment constraints, data modality availability, and application-specific forensic requirements.

Collectively, these findings quantitatively confirm that performance gains from multimodal fusion must be balanced against constraints on interpretability, privacy preservation, and regulatory compliance, which are particularly acute in healthcare cyber-forensic systems.

3.6 Multimodal fusion as a cyber-forensic system architecture

Figure 5 illustrates the proposed layered, multimodal cyber-forensic system architecture. The system comprises five sequential layers, each representing a distinct functional component of the forensic pipeline.

The input layer aggregates heterogeneous data sources, including network logs, IoMT device telemetry, EHR access logs, and medical sensor data. These diverse modalities reflect the complex cyber-physical nature of healthcare systems, where attacks may span multiple domains simultaneously [52, 53].

The preprocessing layer ensures data consistency and interoperability through cleaning, temporal and semantic alignment, and feature extraction. This preprocessing is critical in multimodal systems because heterogeneous healthcare data often differ in structure, scale, and timing, necessitating harmonization before integration [54, 55].

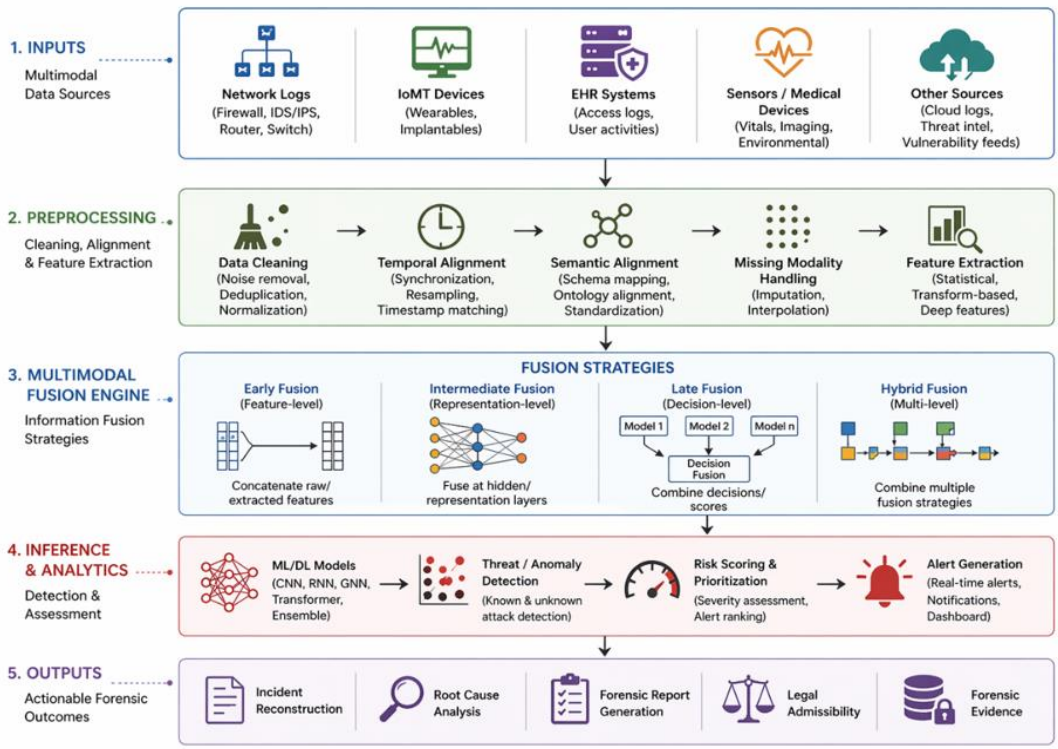

Figure 5. Layered multimodal cyber-forensic system architecture for healthcare cybersecurity

The multimodal fusion engine is the core component of the system, integrating multiple data modalities through early, intermediate, late, and hybrid fusion strategies. These approaches enable the system to capture complementary information across modalities, improving detection robustness and accuracy [56]. From a system perspective, the fusion engine serves as a central processing unit, transforming distributed data streams into unified representations for downstream analysis.

The inference and analytics layer apply machine learning and deep learning models to detect intrusions and anomalies, thereby identifying cyber threats. Prior studies have shown that integrating multiple data sources significantly improves detection performance compared with single-modality approaches, especially in complex cybersecurity environments [56]. This layer also supports risk scoring, prioritization, and real-time alert generation, enabling operational decision-making.

Lastly, the output layer translates analytical results into actionable forensic outcomes, such as incident reconstruction, root cause analysis, forensic reporting, and legal admissibility. This stage ensures that system outputs are not only accurate but also interpretable and suitable for forensic investigation and regulatory compliance [57].

This concludes that the architecture demonstrates that multimodal fusion should be conceptualized as a system-level capability in which each layer contributes to detection accuracy, operational reliability, and evidentiary validity. This system-oriented design is particularly critical in healthcare cybersecurity, where data heterogeneity, regulatory constraints, and safety-critical operations require integrated and robust cyber-forensic solutions.

4. DISCUSSION

This review finds that multimodal fusion is evolving from a purely performance-oriented machine-learning technique into a broader evidentiary approach capable of supporting cyber-forensic analysis in complex digital environments, such as healthcare systems. Synthesis: Combining complementary data representations yields higher accuracy and robustness than analyzing single sources [28, 48]. This improvement stems from capturing diverse attack signatures, such as spatial patterns in images and temporal patterns in audio or graphs [23, 35].

4.1 Healthcare-oriented forensic evaluation framework

To systematically evaluate multimodal fusion systems for healthcare cyber-forensic readiness, this study introduces a four-step, multidimensional assessment framework. Despite conventional evaluations that focus solely on predictive performance, the framework incorporates forensic, privacy, and operational considerations that are critical to real-world deployment. It evaluates each method across four dimensions: transferability, evidentiary reliability, privacy compliance, and operational suitability. Figure 6 illustrates the multidimensional trade-offs across these dimensions.

Step 1: Transferability Assessment

This step assesses whether a multimodal fusion method developed in a source domain, such as IoT, vehicular networks, or malware detection, can be adapted for use in healthcare cyber-forensic environments. The assessment is based on three criteria:

i. C1.1 Cross-domain validation: Whether the method has been tested across multiple datasets or domains?
ii. C1.2 Data modality alignment: Do the fused modalities have structural equivalents in healthcare systems (such as network logs ↔ EHR access logs, IoT ↔ IoMT)?
iii. C1.3 Architecture generalizability: Can the fusion architecture incorporate new modalities without a complete redesign?

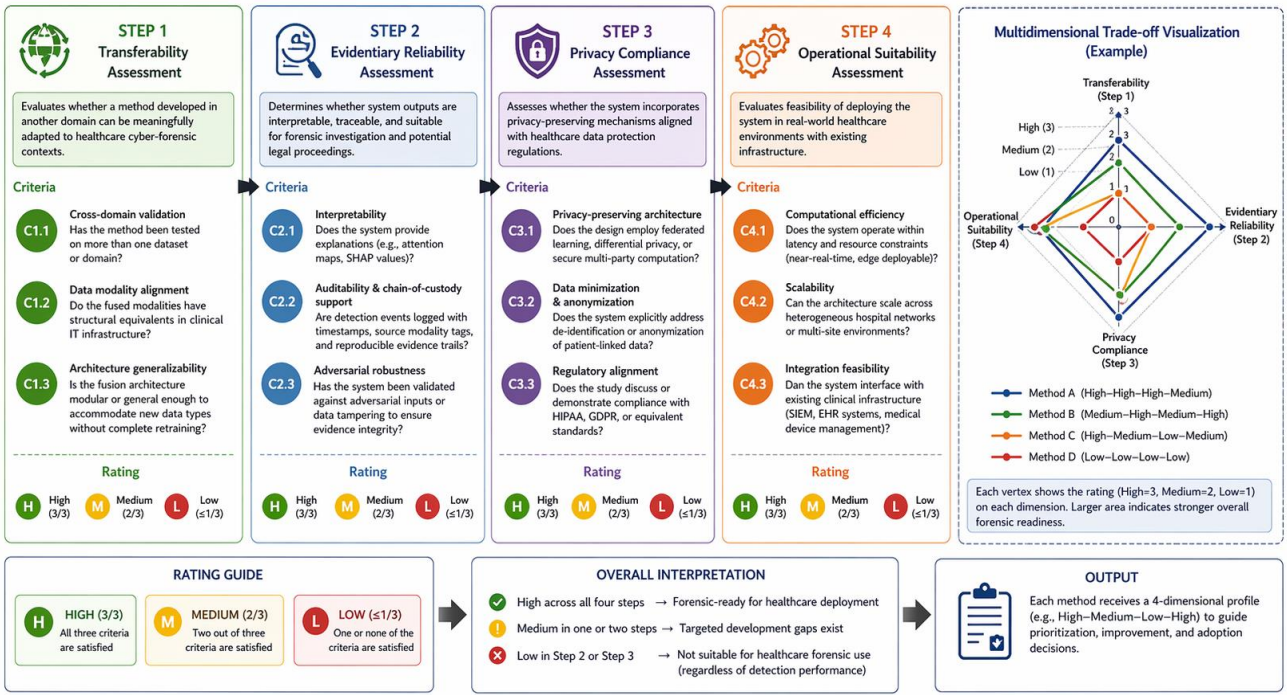

Figure 6. Multidimensional forensic evaluation framework for multimodal fusion methods in healthcare cyber forensics

Step 2: Evidentiary Reliability Assessment

This step determines whether system outputs are suitable for forensic investigation and for legal scrutiny. The criteria include:

i. C2.1 Interpretability: Availability of explainability mechanisms such as attention maps, SHAP values.
ii. C2.2 Auditability: Presence of traceable logs, timestamps, and reproducible evidence trails.
iii. C2.3 Adversarial robustness: Validation against adversarial manipulation or data tampering.

Step 3: Privacy Compliance Assessment

This step assesses alignment with healthcare data protection regulations. The criteria include:

i. C3.1 Privacy-preserving architecture: Use of federated learning, differential privacy, or secure computation.
ii. C3.2 Data anonymization: Explicit de-identification or minimization strategies.
iii. C3.3 Regulatory alignment: Compliance with HIPAA, GDPR, or equivalent standards.

Step 4: Operational Suitability Assessment

This step assesses real-world deployability in healthcare environments. The criteria include:

i. C4.1 Computational efficiency: Ability to operate under latency and resource constraints
ii. C4.2 Scalability: Capability to function across multi-site hospital systems
iii. C4.3 Integration feasibility: Compatibility with clinical systems (e.g., SIEM, EHR platforms)

Applying these four steps produces a multidimensional rating profile (e.g., High–Medium–Low) that enables systematic comparison across studies. A system rated High across all dimensions is considered forensic-ready for healthcare deployment. Methods rated medium highlight areas for targeted improvement, whereas those rated low in evidentiary reliability or privacy compliance are considered unsuitable for healthcare forensic applications, regardless of detection performance.

The results display that while many methods perform well in controlled environments, fewer are fully compliant with privacy standards and ready for operational use. This comprehensive review emphasizes important gaps in the application of multimodal fusion techniques in real-world forensic healthcare. In healthcare and related fields, forensic analysts should use all available data, such as combining EHR logs with network traffic or medical image metadata with patient records, when reconstructing incidents. Figure 6 visually represents this evaluation framework by mapping multimodal fusion methods across the four assessment dimensions. Each axis represents one dimension of transferability, evidentiary reliability, privacy compliance, and operational suitability, enabling readers to compare strengths and weaknesses across methods. This multidimensional assessment highlights the inherent trade-offs among detection performance, interpretability, privacy preservation, and deployment feasibility in healthcare settings.

4.1.1 Framework application procedure and composite Forensic Readiness Score

To enhance the practical applicability and reproducibility of the proposed forensic evaluation framework, this subsection defines a structured scoring procedure that operationalizes the four-step assessment into a quantitative Composite Forensic Readiness Score (FRS). This extension transforms the framework from a conceptual model into a replicable evaluation method that can be consistently applied across studies and deployment contexts.

Step 1: Numerical Mapping of Qualitative Ratings

Each qualitative dimension rating produced by Steps 1- 4 is mapped to a numerical score to enable standardized, cross-study comparison:

High = 3 (the method fully satisfies the criteria for this dimension)

Medium = 2 (the method partially satisfies the criteria, with identifiable gaps)

Low = 1 (the method does not satisfy the criteria, or the evidence is absent)

This mapping ensures consistency when comparing multimodal fusion methods across different studies and application domains.

Step 2: Dimension-Level Scoring

For each study, evaluators independently assign a rating (High / Medium / Low) for each of the four framework dimensions based on the extent to which the study satisfies the corresponding sub-criteria defined in Steps 1- 4:

• Transferability (T) - assessed via C1.1- C1.3
• Evidentiary Reliability (R) - assessed via C2.1- C2.3
• Privacy Compliance (P) - assessed via C3.1- C3.3
• Operational Suitability (O) - assessed via C4.1- C4.3

These ratings are then mapped to scores (1–3) using Step A, producing a structured rating profile (e.g., High-Medium–Low-High) and a corresponding numeric profile for each evaluated method.

Step 3: Composite FRS

The four-dimensional scores are aggregated into a single Composite FRS using a weighted sum:

$$FRS = w_1T + w_2R + w_3P + w_4O$$

where, T, R, P, and O are the numerical scores (1–3) for each dimension, and w_1–w_4 are dimension weights summing to 1.0 ($\Sigma w_i = 1.0$). By default, equal weights are applied ($w_i = 0.25$ each), yielding an unweighted FRS in the range 1.0 - 3.0. However, weights may be adjusted to reflect application-specific priorities. In healthcare cyber-forensic contexts, greater weight should be assigned to evidentiary reliability and privacy compliance due to regulatory and legal requirements, for example, $w_2 = 0.35$ (Evidentiary Reliability) and $w_3 = 0.30$ (Privacy Compliance), with $w_1 = 0.20$ (Transferability) and $w_4 = 0.15$ (Operational Suitability). Table 9 presents the FRS interpretation tiers applicable under both weighting schemes.

Step 4: Interpretation of the FRS

The resulting FRS is interpreted using the following readiness tiers, which apply regardless of the weighting scheme used.

Regardless of the overall FRS, a rating of Low on either Evidentiary Reliability (R = 1) or Privacy Compliance (P = 1) automatically disqualifies the method from healthcare forensic deployment, reflecting the non-negotiable nature of chain-of-custody integrity and HIPAA/GDPR compliance in clinical environments. In such cases, the method is classified as Not Ready, irrespective of its FRS value.

This structured procedure ensures that the evaluation of multimodal fusion systems extends beyond detection performance to include forensic validity, regulatory compliance, and the feasibility of real-world clinical

deployment. While the FRS provides a useful quantitative summary for cross-study comparison, it should always be interpreted alongside the qualitative dimension profiles to account for context-specific operational requirements and the threshold constraints described above.

The results show that while many methods perform well in controlled environments, fewer are fully compliant with privacy standards and ready for operational use. This comprehensive review highlights significant gaps in the application of multimodal fusion techniques in real-world forensic healthcare. In healthcare and related fields, forensic analysts should use all available data, such as combining EHR logs with network traffic or medical image metadata with patient records, when reconstructing incidents. Figure 7 supports this idea by showing multidimensional trade-offs among the methods being evaluated.

Table 10 provides a structured evaluation of the included multimodal fusion studies using the proposed healthcare-focused forensic framework. Each study is assessed across four main areas: transferability, evidentiary reliability, privacy compliance, and operational suitability. Transferability is rated based on how well methods are validated across different datasets or domains, especially their potential use in healthcare settings. Evidentiary reliability demonstrates how well models generate interpretable, traceable, and audit-ready outputs suitable for forensic investigation. The step addresses a core evaluation dimension, provides explicit criteria and guiding questions, and yields a qualitative rating (High / Medium / Low) that informs the overall assessment. Privacy compliance assesses whether privacy-preserving mechanisms, such as federated learning, differential privacy, or secure data handling practices, meet regulatory standards. Operational suitability assesses the feasibility of deployment in real-world scenarios, considering factors such as computational efficiency, scalability, latency, and integration with existing systems.

Each dimension is qualitatively assessed as High, Medium, or Low based on the model's generalizability, interpretability, privacy-preserving mechanisms, and deployment readiness as reported in the respective studies.

The field is undergoing rapid evolution. Federated learning has been adopted to address data privacy and distribution issues in IoT networks [27, 34]. Few-shot and transfer learning are still underexplored in our dataset; however, early research on large pre-trained multimodal models [27]. Leveraging deep self-supervision on cybersecurity data is a promising approach. Graph neural networks and transformers are becoming increasingly prominent, enabling the fusion of complex relational structures of device graphs and attention-based integration. Explainability is also gaining attention, with multi-view attention used to interpret feature importance, demonstrating how fused models can provide actionable insights [47].

Table 9. Composite Forensic Readiness Score (FRS) interpretation tiers and associated actions

Readiness Tier	FRS Range	Interpretation	Action Indicated
Forensic-Ready	FRS ≥ 2.5	Method satisfies all four dimensions at High or near-High level; suitable for healthcare forensic deployment	Can be adopted with standard clinical governance review
Conditionally Ready	FRS 1.8–2.4	Method performs well on some dimensions but has identifiable gaps requiring targeted improvement	Identify the lowest-scoring dimension(s) and address specific sub-criteria before deployment
Not Ready	FRS < 1.8	The method has critical deficiencies, particularly in evidentiary reliability or privacy compliance	Do not deploy in a healthcare forensic context; fundamental redesign required

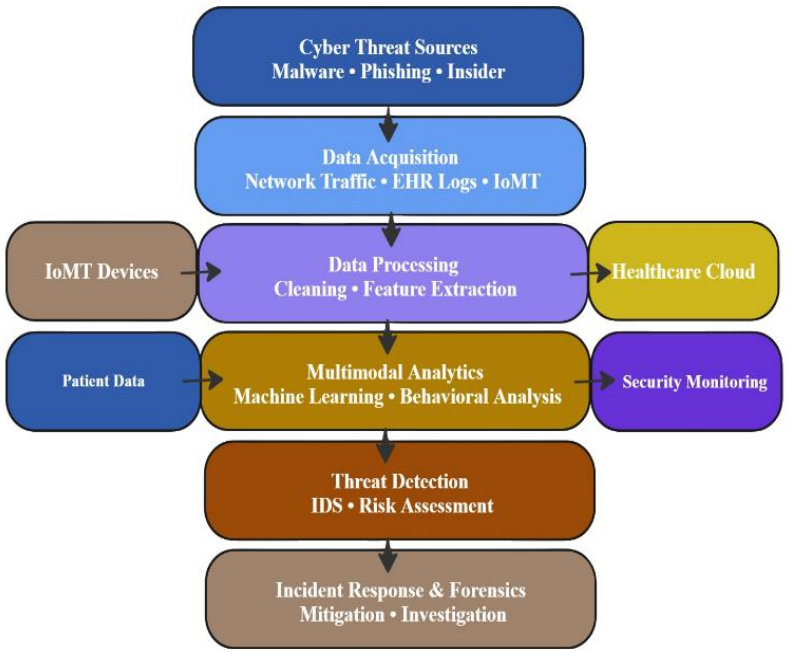

Figure 7. Challenges and constraints of multimodal fusion in healthcare cyber forensics across technical, privacy, and legal dimensions

Table 10. Framework-based evaluation of multimodal fusion studies

Study	Domain	Transferability	Evidentiary Reliability	Privacy Compliance	Operational Suitability	Key Insight
[23]	Malware classification	Medium	High	Low	Medium	Feature fusion improves detection but lacks privacy considerations
[24]	Vehicular IDS	High	High	Low	High	Real-time capable; strong forensic traceability
[25]	Network IDS	High	Medium	Low	High	High accuracy but limited interpretability
[26]	ICS security	High	High	Low	Medium	Cross-domain fusion enhances forensic reliability
[27]	Android malware	Medium	Medium	Low	Medium	Transformer-based fusion with moderate explainability
[29]	Darknet analysis	Medium	Medium	Low	Medium	Graph-based fusion improves contextual threat detection
[30]	Android malware	Medium	High	Low	Medium	Hybrid static–dynamic fusion enhances robustness
[31]	Mobile malware	Medium	High	Low	Medium	Multi-stage fusion improves adaptability under concept drift
[32]	Multimodal learning (survey)	Medium	Medium	Low	Low	Provides a methodological overview but lacks empirical validation
[33]	Malware classification	Medium	Medium	Low	Medium	Bimodal learning improves classification performance
[34]	Vehicular IDS	High	High	Medium	High	Transformer–GNN fusion improves cross-domain detection
[35]	Malware detection	Medium	High	Low	Medium	Ensemble deep learning improves accuracy and robustness
[36]	Malware classification	Medium	Medium	Low	Medium	Structural Entropy enhances malware representation
[37]	Ransomware detection	High	High	Low	High	Cross-modal transformers improve ransomware detection
[38]	Android malware	Medium	Medium	Low	Medium	Hierarchical modeling improves detection accuracy
[39]	Android malware	Medium	High	Low	Medium	Multimodal DL enhances malware classification performance
[40]	Vehicular IDS	High	High	Medium	High	Federated learning supports scalable vehicular IDS
[41]	IoT IDS	High	Medium	Medium	High	Knowledge distillation improves generalization in IoT systems
[42]	IoT IDS	High	Medium	Medium	High	Personalized federated IDS improves adaptability
[43]	Healthcare IDS	Medium	Medium	Medium	Medium	Multimodal IDS enhances healthcare system security
[44]	Network security	High	Medium	Low	High	KAN-based fusion improves enterprise threat detection
[45]	Vehicular IDS	High	High	Medium	High	Federated LSTM enables efficient real-time detection
[46]	IoT IDS	High	High	Medium	High	Knowledge distillation improves robustness and efficiency
[47]	Vehicular IDS	High	Medium	Medium	High	Lightweight models enable edge deployment
[48]	Cyberattack prediction	High	Medium	Low	Medium	Multimodal learning improves predictive threat modeling

Note: Ratings (High, Medium, Low) are determined by each study's overall transferability, interpretability, auditability (evidentiary reliability), privacy-preserving techniques (privacy compliance), and feasibility for real-world deployment (operational suitability). (Rating: High (3/3), Medium (2/3), Low (≤ 1/3)).

Healthcare systems handle sensitive personal data and critical care, so forensic tools must balance detection capabilities with privacy concerns. The success of federated, multi-party approaches here is promising: they enable cross-site model training without sharing raw health data [42, 47]. Policy frameworks (e.g., HIPAA, GDPR) will likely steer forensic analytics toward such privacy-preserving integration. Additionally, combining blockchain or SMPC with multimodal analytics to secure healthcare data could create a multi-layered cyber defense. High-performance multimodal models also underscore the importance of ensuring the trustworthiness of healthcare IT, specifically robustness against adversarial manipulation [58].

4.2 Ethical issues in multimodal forensic fusion

While multimodal fusion improves cyber forensics by increasing accuracy and resilience, it also raises several ethical and legal concerns, especially in the healthcare sector. The studies reviewed indicate that combining diverse data sources, such as network logs, device telemetry, and patient records, can unintentionally expose sensitive information and heighten privacy risks [7, 11, 30]. Healthcare data are particularly vulnerable because integrating across modalities often requires access to personally identifiable and clinical information [13]. Without proper anonymization, such integration could contravene data protection standards such as

HIPAA and GDPR [18, 58]. In healthcare environments, ethical compliance and forensic reliability are closely interconnected. Systems that produce highly accurate predictions but lack transparency or auditability may struggle to meet legal or institutional standards required for incident investigation and regulatory reporting.

Furthermore, bias and fairness issues arise in algorithmic fusion models. Research on multimodal intrusion detection and malware detection seldom evaluates model bias or disparities in representation across datasets, which can result in unfair outcomes or incorrect classification of benign user activity [8, 27, 35]. Transparency and explainability are also lacking in the deep fusion frameworks. Combining hybrid CNN-LSTM and transformer architectures acts as black boxes, making legal admissibility and accountability in forensic settings more difficult contexts [9, 10].

Finally, combining distributed data through federated or cloud-based learning raises additional ethical concerns regarding consent, ownership, and data sovereignty [27, 42, 46]. While such methods preserve privacy by avoiding the exchange of raw data, they may still leak metadata or learned representations. The reviewed studies emphasize the need for privacy-preserving fusion frameworks, improved transparency, and standardized ethical evaluation.

4.3 Strengths and limitations

A key strength of this review is its translational approach, linking multimodal cybersecurity research across various fields with the specific needs of healthcare cyber-forensic practice. However, the diversity of datasets, evaluation metrics, and experimental conditions across studies limits the ability to conduct quantitative meta-analysis [6].

4.4 Policy and research implications

Findings focus on actionable steps: (1) Develop benchmark datasets for healthcare cyber data that simulate multimodal cyber-attacks, like CIC or DARPA datasets used in general networks, to evaluate fusion techniques [25, 26]. (2) Promote interdisciplinary collaboration by involving security experts and healthcare practitioners to identify relevant modalities, such as incorporating patient monitoring data into threat assessment analysis [6, 16]. (3) Revise privacy regulations and cyber forensic guidelines to support federated and privacy-preserving fusion methods, ensuring legal frameworks enable cross-institution threat analysis without compromising patient data [52, 58]. (4) Future research should focus on finding ways to address new malware or attacks and developing deep explainability for fused models to enhance transparency in forensic analysis [32]. This shift from performance-focused models to forensic-ready systems reflects the growing need for cybersecurity solutions that are not only accurate but also interpretable and legally defensible.

5. CONCLUSION

This systematic review shows that multimodal fusion architectures are now essential in digital forensics, enabling the integration of sensor inputs, logs, network traces, and contextual metadata into cohesive evidential representations. Early efforts focused on simple feature concatenation across modalities. However, modern methods employ hierarchical encoders, attention mechanisms, and graph-based models to learn cross-modal relationships and handle missing or noisy data more effectively. These architectures improve the sensitivity and specificity of forensic tools by reducing false positives and enabling more accurate attribution. Despite these advances, few studies thoroughly assess the trade-offs among early-, intermediate-, and late-fusion strategies or compare their performance across domains.

The review's second question addressed application areas and transferability. Most research on multimodal fusion targets digital crime scene reconstruction, network intrusion detection, and behavioral biometrics. Emerging work also examines toxicology and IoT surveillance. Although examples of healthcare applications are limited, the report shows that algorithms designed for cybercrime and industrial control systems can be adapted for healthcare cybersecurity, mainly within IoMT and clinical IT networks. In these areas, various signals, such as device logs, physiological data, and access records, can help quickly detect anomalies following incidents. However, using these methods presents challenges, including privacy concerns, regulatory requirements, and life-critical implications, which demand robust, interpretable models.

The third question pointed out ongoing challenges. Researchers must address data heterogeneity, inconsistent standards, limited labeled datasets, and the risk of missing modalities. Privacy, ethical, and legal concerns often take a backseat to technical performance, and few studies focus on adversarial robustness, explainability, and chain-of-custody factors that are crucial for courtroom use. Therefore, this review urges future research into privacy-preserving and legally compliant fusion models that establish clear evidence trails, safeguard sensitive patient data, and withstand adversarial attacks.

To advance cyber forensic science and healthcare security, collaboration across various disciplines is essential. Cybersecurity experts, healthcare technologists, forensic specialists, and policymakers should work together to set standards, establish data-sharing protocols, and develop regulations that encourage innovation while safeguarding individual rights. In conclusion, the future of healthcare cybersecurity may rely on multimodal forensic systems that combine different digital evidence, enabling investigators to detect, interpret, and reconstruct cyber incidents within increasingly complex clinical information systems. Future work should focus on validating multimodal forensic systems in real healthcare settings and establishing standards for evidentiary reliability and regulatory compliance.

STATEMENT ON THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE

Generative artificial intelligence (AI) tools (e.g., ChatGPT) were used solely to assist with language editing, grammar correction, and improving the manuscript's clarity and readability. The research design, methodology, literature review, data extraction, analysis, interpretation of results, and all scientific contributions were conducted entirely by the authors. The authors take full responsibility for the accuracy, integrity, and final content of this manuscript.

REFERENCES

[1] Kruse, C.S., Frederick, B., Jacobson, T., Monticone,

D.K. (2017). Cybersecurity in healthcare: A systematic review of modern threats and trends. Technology and Health Care, 25(1): 1-10. https://doi.org/10.3233/THC-161263

[2] Ghafur, S., Kristensen, S., Honeyford, K., Martin, G., Darzi, A., Aylin, P. (2019). A retrospective impact analysis of the WannaCry cyberattack on the NHS. npj Digital Medicine, 2: 98. https://doi.org/10.1038/s41746-019-0161-6

[3] Portela, D., Nogueira-Leite, D., Almeida, R., Cruz-Correia, R. (2023). Economic impact of a hospital Cyberattack in a national health system: Descriptive case study. JMIR Formative Research, 7: e41738. https://doi.org/10.2196/41738

[4] Kraetzer, C., Makrushin, A., Dittmann, J., Hildebrandt, M. (2021). Potential advantages and limitations of using Information fusion in media forensics—A discussion on the example of detecting face morphing attacks. EURASIP Journal on Information Security, 2021: 9. https://doi.org/10.1186/s13635-021-00123-4

[5] Hu, X.Y., Lu, D.H., Liu, Y., Wu, B., Yang, F. (2026). APGFusion: Adaptive PoolFormer and CNN medical image fusion network based on convolutional gated linear units. Expert Systems with Applications, 307: 130867. https://doi.org/10.1016/j.eswa.2025.130867

[6] Chen, Q.Q., Li, J.P., Haq, A.U., Agbley, B.L.Y., Hussain, A., Khan, I., Khan, R.U., Khan, J., Ali, I. (2023). A multimodal network security framework for healthcare based on deep learning. Computational Intelligence and Neuroscience, 2023(1): 9041355. https://doi.org/10.1155/2023/9041355

[7] Teoh, J.R., Dong, J., Zuo, X.W., Lai, K.W., Hasikin, K., Wu, X. (2024). Advancing healthcare through multimodal data fusion: A comprehensive review of techniques and applications. PeerJ Computer Science, 10: e2298. https://doi.org/10.7717/peerj-cs.2298

[8] Xu, C.Y., Zhan, Y., Wang, Z.Q., Yang, J. (2025). Multimodal fusion-based few-shot network intrusion detection system. Scientific Reports, 15: 21986. https://doi.org/10.1038/s41598-025-05217-4

[9] Swofford, H. (2024). Forensic Science Environmental Scan 2023. NIST Interagency Report, IR 8515. https://doi.org/10.6028/NIST.IR.8515

[10] Scurich, N., Faigman, D.L., Albright, T.D. (2023). Scientific guidelines for evaluating the validity of forensic feature-comparison methods. Proceedings of the National Academy of Sciences, 120(41): e2301843120. https://doi.org/10.1073/pnas.2301843120

[11] Tang, Z.Y., Tang, Z., Wu, J. (2025). Anomaly detection in medical via multimodal foundation models. Frontiers in Bioengineering and Biotechnology, 13: 1644697. https://doi.org/10.3389/fbioe.2025.1644697

[12] Krones, F., Marikkar, U., Parsons, G., Szmul, A., Mahdi, A. (2025). Review of multimodal machine learning approaches in healthcare. Information Fusion, 114: 102690. https://doi.org/10.1016/j.inffus.2024.102690

[13] Kline, A., Wang, H.Y., Li, Y.K., Dennis, S., Hutch, M., Xu, Z.X., Wang, F., Cheng, F.X., Luo, Y. (2022). Multimodal machine learning in precision health: A scoping review. npj Digital Medicine, 5: 171. https://doi.org/10.1038/s41746-022-00712-8

[14] Ahmed, S.F., Shawon, S.S., Bhuyian, A., Afrin, S., Mehjabin, A., Kuldeep, S.A., Alam, M.S.B., Gandomi, A.H. (2025). Forensics and security issues in the Internet of Things. Wireless Networks, 31: 3431-3466. https://doi.org/10.1007/s11276-025-03942-2

[15] Akinbi, A., MacDermott, Á., Ismael, A.M. (2022). A systematic literature review of blockchain-based Internet of Things (IoT) forensic investigation process models. Forensic Science International: Digital Investigation, 42-43: 301470. https://doi.org/10.1016/j.fsidi.2022.301470

[16] Shaik, T., Tao, X.H., Li, L., Xie, H.R., Velásquez, J.D. (2024). A survey of multimodal information fusion for smart healthcare: Mapping the journey from data to wisdom. Information Fusion, 102: 102040. https://doi.org/10.1016/j.inffus.2023.102040

[17] Zhang, Y., Sheng, M., Liu, X.Y., Wang, R.Y., Lin, W.H., Ren, P., Wang, X., Zhao, E.L., Song, W.C. (2022). A heterogeneous multimodal medical data fusion framework supporting hybrid data exploration. Health Information Science and Systems, 10: 22. https://doi.org/10.1007/s13755-022-00183-x

[18] Kruse, C., Heinemann, K. (2022). Facilitators and barriers to telemedicine adoption of telemedicine during the first year of COVID-19: Systematic review. Journal of Medical Internet Research, 24(1): e31752. https://doi.org/10.2196/31752

[19] Ahmed, A.A., Farhan, K., Jabbar, W.A., Al-Othmani, A., Abdulrahman, A.G. (2024). IoT forensics: Current perspectives and future directions. Sensors, 24(16): 5210. https://doi.org/10.3390/s24165210

[20] Sun, H., Wan, L., Liu, M., Wang, B. (2023). Few-shot network intrusion detection based on prototypical capsule network with attention mechanism. PLoS One, 18(4): e0284632. https://doi.org/10.1371/journal.pone.0284632

[21] Tian, J.Y., Wang, Z.M., Fang, H., Chen, L.M., Qin, J., Chen, J., Wang, Z.H. (2022). Few-shot learning-based network intrusion detection through an enhanced parallelized triplet network. Security and Communication Networks, 2022(1): 3317048. https://doi.org/10.1155/2022/3317048

[22] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372: n21. https://doi.org/10.1136/bmj.n71

[23] Gibert, D., Planes, J., Mateu, C., Le, Q. (2022). Fusing feature engineering and deep learning: A case study for malware classification. Expert Systems with Applications, 207: 117957. https://doi.org/10.1016/j.eswa.2022.117957

[24] Kang, H., Vo, T., Kim, H.K., Hong, J.B. (2024). CANival: A multimodal approach to intrusion detection on the vehicle CAN bus. Vehicular Communications, 50: 100845. https://doi.org/10.1016/j.vehcom.2024.100845

[25] Shi, S., Han, D., Cui, M. (2023). A multimodal hybrid parallel network intrusion detection model. Connection Science, 35(1): 2227780. https://doi.org/10.1080/09540091.2023.2227780

[26] Sahu, A., Mao, Z., Wlazlo, P., Huang, H., Davis, K., Goulart, A. (2021). Multi-source multi-domain data fusion for cyberattack detection in power systems. IEEE Access, 9: 119118-119138. https://doi.org/10.1109/ACCESS.2021.3106873

[27] Li, X., Liu, L., Liu, Y.Z., Zhao, Y., Zhang, P., Liu, H.X. (2025). Multimodal fusion for Android malware detection based on large pre-trained models. IEEE Transactions on Software Engineering, 51(5): 1569-

1590. https://doi.org/10.1109/TSE.2025.3557577
[28] Ismail, S.J.I., Hendrawan, Rahardjo, B., Juhana, T., Musashi, Y. (2025). MIDALF—Multimodal image and audio late fusion for malware detection. EURASIP Journal on Information Security, 2025: 5. https://doi.org/10.1186/s13635-025-00188-5
[29] Hwang, Y., Kurt, F., Curebal, F., Keskin, O., Subasi, A. (2026). ContextualGraph-LLM: A multimodal framework for enhanced Darknet traffic analysis. Expert Systems with Applications, 297: 129298. https://doi.org/10.1016/j.eswa.2025.129298
[30] Kauser.Sk, H., Anu.V, M. (2025). Hybrid deep learning model for accurate and efficient Android malware detection using DBN-GRU. PLoS One, 20(5): e0310230. https://doi.org/10.1371/journal.pone.0310230
[31] Augello, A., De Paola, A., Lo Re, G. (2025). Hybrid multilevel detection of mobile devices malware under concept drift. Journal of Network and Systems Management, 33: 36. https://doi.org/10.1007/s10922-025-09906-3
[32] Zhao, F., Zhang, C.C., Geng, B.C. (2024). Deep multimodal data fusion: A survey. ACM Computing Surveys, 56(9): 1-36. https://doi.org/10.1145/3649447
[33] Gibert, D., Mateu, C., Planes, J. (2020). Orthrus: A bimodal learning architecture for malware classification. In 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, pp. 1-8. https://doi.org/10.1109/IJCNN48605.2020.9206671
[34] Yonpang, S., Sahba, B.C., Ngueusseu, H., Kombou, V., Leaticia, K.S. (2024). MM-IDS: A multi-modal intrusion detection system for enhanced vehicular network cybersecurity. SSRN. http://doi.org/10.2139/ssrn.5223388
[35] Nazim, S., Alam, M.M., Rizvi, S., Mustapha, J.C., Hussain, S.S., Su'ud, M.M. (2025). Multimodal malware classification using proposed ensemble deep neural network framework. Scientific Reports, 15: 18006. https://doi.org/10.1038/s41598-025-96203-3
[36] Gibert, D., Mateu, C., Planes, J., Vicens, R. (2018). Classification of malware by using structural entropy on convolutional neural networks. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1): 7759-7764. https://doi.org/10.1609/aaai.v32i1.11409
[37] Alzahrani, S., Xiao, Y., Asiri, S., Alasmari, N., Li, T. (2025). RansomFormer: A cross-modal transformer architecture for ransomware detection via the fusion of byte and API features. Electronics, 14(7): 1245. https://doi.org/10.3390/electronics14071245
[38] Chen, H., Li, Z.Q., Jiang, Q.S., Rasool, A., Chen, L.F. (2021). A hierarchical approach for Android malware detection using authorization-sensitive features. Electronics, 10(4): 432. https://doi.org/10.3390/electronics10040432
[39] Arrowsmith, J., Susnjak, T., Jang-Jaccard, J. (2025). Multimodal deep learning for Android malware classification. Machine Learning and Knowledge Extraction, 7(1): 23. https://doi.org/10.3390/make7010023
[40] Althunayyan, M., Javed, A., Rana, O. (2024). A robust multi-stage intrusion detection system for in-vehicle network security using hierarchical federated learning. Vehicular Communications, 49: 100837. https://doi.org/10.1016/j.vehcom.2024.100837
[41] Shen, J.Y., Yang, W.Z., Chu, Z.W., Fan, J.N., Niyato, D., Lam, K.Y. (2024). Effective intrusion detection in heterogeneous Internet-of-Things networks via ensemble knowledge distillation-based federated learning. In ICC 2024 - IEEE International Conference on Communications, Denver, CO, USA, pp. 2034-2039. https://doi.org/10.1109/ICC51166.2024.10622262
[42] Singh, G., Sood, K., Rajalakshmi, P., Xiang, Y. (2026). Sentinel: Dynamic knowledge distillation for personalized federated intrusion detection in heterogeneous IoT networks. IEEE Internet of Things Journal, 13(7): 14682-14694. https://doi.org/10.1109/JIOT.2026.3650848
[43] Nguyen, P.T., Huynh, V.D.B., Vo, K.D., Phan, P.T., Elhoseny, M., Le, D.N. (2021). Deep learning-based optimal multimodal fusion framework for intrusion detection systems for healthcare data. Computers, Materials & Continua, 66(3): 2556-2571. https://doi.org/10.32604/cmc.2021.012941
[44] Hu, Z., Wang, L., Ding, X., Zhao, L., Xue, M. (2025). Multimodal data fusion defense strategy for campus network security: Research on Kolmogorov–Arnold networks combined with B-spline function. Discover Computing, 28: 83. https://doi.org/10.1007/s10791-025-09593-3
[45] Martínez, M.Z., Marin-Perez, R., Gomez, A.F.S. (2025). Development of an in-vehicle intrusion detection model integrating federated learning and LSTM networks. Information, 16(4): 292. https://doi.org/10.3390/info16040292
[46] Quyen, N.H., Duy, P.T., Nguyen, N.T., Khoa, N.H., Pham, V.H. (2025). FedKD-IDS: A robust intrusion detection system using knowledge distillation-based semi-supervised federated learning and anti-poisoning attack mechanism. Information Fusion, 117: 102807. https://doi.org/10.1016/j.inffus.2024.102807
[47] Li, J.J., Ma, Y.Y., Bai, J.H., Chen, C.M., Xu, T.T., Ding, C. (2025). A lightweight intrusion detection system with dynamic feature fusion federated learning for vehicular network security. Sensors, 25(15): 4622. https://doi.org/10.3390/s25154622
[48] Dong, J.P., Hao, M.M., Ding, F.Y., Chen, S., Wu, J.J., Zhuo, J., Jiang, D. (2025). A novel multimodal data fusion framework: Enhancing prediction and understanding of inter-state cyberattacks. Big Data and Cognitive Computing, 9(3): 63. https://doi.org/10.3390/bdcc9030063
[49] Gibert, D., Mateu, C., Planes, J. (2020). The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. Journal of Network and Computer Applications, 153: 102526. https://doi.org/10.1016/j.jnca.2019.102526
[50] Atrey, P.K., Hossain, M.A., El Saddik, A., Kankanhalli, M.S. (2010). Multimodal fusion for multimedia analysis: A survey. Multimedia Systems, 16: 345-379. https://doi.org/10.1007/s00530-010-0182-0
[51] Aksu, F., Gelardi, F., Chiti, A., Soda, P. (2025). Multi-stage intermediate fusion for multimodal learning to classify non-small cell lung cancer subtypes from CT and PET. Pattern Recognition Letters, 193: 86-93. https://doi.org/10.1016/j.patrec.2025.04.001
[52] Jiang, L.J., Li, Q.M., Che, X., Chen, X. (2025). A knowledge distillation enhanced semi-supervised federated learning framework for intrusion detection in EV charging networks. IEEE Internet of Things Journal,

12(16): 34360-34373. https://doi.org/10.1109/JIOT.2025.3577666
[53] Ikegwu, A.C., Alo, U.R., Nweke, H.F. (2025). Cyber threats in mobile healthcare applications: Systematic review of enabling technologies, threat models, detection approaches, and future directions. Discover Computing, 28: 152. https://doi.org/10.1007/s10791-025-09686-z
[54] Huang, S.C., Pareek, A., Seyyedi, S., Banerjee, I., Lungren, M.P. (2020). Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines. NPJ Digital Medicine, 3: 136. https://doi.org/10.1038/s41746-020-00341-z
[55] Baltrušaitis, T., Ahuja, C., Morency, L.P. (2019). Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2): 423-443. https://doi.org/10.1109/TPAMI.2018.2798607
[56] Buczak, A.L., Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2): 1153-1176. https://doi.org/10.1109/COMST.2015.2494502
[57] Casey, E. (2011). Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet (3rd ed.). Academic Press. https://dl.acm.org/doi/book/10.5555/2021194.
[58] Chevrier, R., Foufi, V., Gaudet-Blavignac, C., Robert, A., Lovis, C. (2019). Use and understanding of anonymization and de-identification in the biomedical literature: Scoping review. Journal of Medical Internet Research, 21(5): e13484. https://doi.org/10.2196/13484