REVIEW **Open access**

Distribution Shift in Healthcare AI: Detection Methods, Adaptation Strategies, and Failure Taxonomies

Anna Kowalska[1*], Piotr Nowak[1], Tomasz Zielinski[2], Katarzyna Mazur[1]

Abstract

Distribution shifts pose a major challenge for artificial intelligence (AI) deployed in healthcare systems, as models trained on historical or controlled datasets often perform poorly in evolving clinical environments. This narrative review synthesizes current approaches for detecting, adapting to, and classifying failures associated with distribution shifts in AI-driven healthcare analytics. Sources of shift—including changes in patient demographics, imaging protocols, institutional practices, and temporal dynamics—can significantly affect clinical decision support, predictive modeling, and operational analytics.

We examine detection strategies based on statistical divergence monitoring and discuss adaptation methods such as domain adaptation and privacy-preserving learning approaches designed to maintain model performance across institutions. Failure modes are organized into core categories, including covariate shift, label shift, and concept drift, with particular attention to healthcare-specific risks such as bias amplification and breakdowns in continuous monitoring systems.

From a systems perspective, the review highlights the importance of integrating shift detection with clinical analytics pipelines, governance mechanisms, and explainable AI tools to support safe deployment. We propose an interpretive framework linking data ingestion, model inference, intervention feedback, and oversight processes within healthcare infrastructures. Despite advances in detection and adaptation techniques, real-time operational deployment and standardized failure classification remain significant gaps. Strengthening these areas is essential for developing resilient AI systems capable of maintaining reliability in dynamic healthcare environments.

Keywords Clinical analytics, Healthcare AI, Domain adaptation, Distribution shift, Dataset shift detection, Failure taxonomies

*Correspondence:
Anna Kowalska
anna.kowalska@gmail.com
[1] Department of Medical Informatics, Faculty of Medicine, University of Warsaw, Warsaw, Poland
[2] Department of Digital Systems Engineering, Faculty of Engineering, Warsaw University of Technology, Warsaw, Poland

Introduction

The emergence of AI in healthcare systems and the challenge of distribution shifts

The integration of artificial intelligence (AI) into healthcare systems has transformed clinical analytics from static data processing to dynamic, predictive infrastructures capable of supporting real-time decision-making. Over the past decade, AI models have been deployed across various healthcare layers, including diagnostic imaging, risk stratification, and resource allocation, leveraging vast datasets from electronic health records (EHRs) and multimodal sources. However, this rapid adoption has exposed a fundamental vulnerability: distribution shifts, where the statistical properties of deployment data diverge from training distributions, compromising model reliability [1-4]. In healthcare contexts, these shifts arise from heterogeneous factors such as evolving patient populations, changes in clinical protocols, or variations in data acquisition devices, leading to suboptimal performance in predictive analytics and operational systems [2]. For instance, models trained on data from urban tertiary centers may falter when applied to rural settings due to demographic and comorbidity differences, illustrating the systemic implications for equitable healthcare delivery. This phenomenon not only affects individual model outputs but also propagates through interconnected healthcare workflows, potentially amplifying errors in closed-loop systems where AI informs interventions and feedback loops. The literature increasingly recognizes that addressing distribution shifts requires a shift from isolated algorithmic solutions to holistic systems engineering, incorporating detection, adaptation, and failure mitigation at the infrastructure level [3-7]. Comparative analyses across studies reveal that while early AI applications focused on accuracy in controlled environments, contemporary deployments emphasize robustness against real-world variabilities, integrating AI with clinical decision support tools to enhance system resilience [8-20].

Healthcare systems, characterized by their complexity and interconnectivity, demand AI solutions that align with clinical workflows rather than operate in silos. Distribution shifts exacerbate this challenge by introducing uncertainties that can disrupt analytics pipelines, from data ingestion to outcome prediction [16]. Cross-study syntheses indicate that shifts are particularly pronounced in time-sensitive applications, such as intensive care unit (ICU) monitoring, where temporal drifts in patient vital signs can lead to false alarms or missed detections [20-26]. Methodologically, researchers have employed divergence metrics to quantify these shifts, but their integration into scalable healthcare infrastructures remains inconsistent, often limited by computational overhead or lack of interpretability [4, 6]. Systems-level interpretations suggest that AI must evolve towards adaptive architectures that recalibrate in response to detected shifts, thereby supporting continuous improvement in clinical analytics [9, 13]. Moreover, governance frameworks are essential to oversee these adaptations, ensuring alignment with ethical standards and regulatory compliance in diverse healthcare settings [23, 25]. This evolving landscape underscores the need for a comprehensive review that synthesizes detection methods, adaptation strategies, and failure taxonomies through a systems-analytic prism, highlighting how AI can be embedded within healthcare ecosystems to mitigate risks and optimize outcomes.

Types and sources of distribution shifts in clinical data ecosystems

Distribution shifts in healthcare AI manifest in multiple forms, including covariate shift, where input distributions change. At the same time, label relationships remain constant, and concept drift, where underlying data-generating processes evolve [1, 11]. In clinical analytics, covariate shifts often stem from institutional variations, such as differences in imaging equipment or EHR coding practices, leading to degraded model generalization across sites [3, 10]. Label shifts, conversely, occur when disease prevalence alters due to epidemiological changes or diagnostic criteria updates, impacting risk modeling in population health systems [22, 26]. Temporal shifts are prevalent in longitudinal datasets, as seen in EHR-based predictive analytics, where patient management protocols shift over years, necessitating continual model monitoring [9, 27, 28]. Cross-study comparisons reveal that multimodal data integration exacerbates these issues, with shifts in one modality (e.g., radiology images) cascading to affect fused predictions in clinical decision systems [8, 15]. Interpretively, these shifts challenge the foundational assumptions of AI training paradigms, requiring healthcare infrastructures to incorporate robust data pipelines that detect and normalize discrepancies at the ingestion stage [16, 27]. Systems implications extend to operational analytics, where unaddressed shifts can lead to inefficient resource allocation, such as over- or under-prediction of hospital admissions [5, 20].

Further analysis shows that sources of shifts are intertwined with healthcare system dynamics, including demographic transitions, policy changes, and technological upgrades [7, 18]. For example, federated learning approaches aim to mitigate shifts by training across distributed datasets, but privacy constraints limit their efficacy in detecting subtle institutional biases [5, 25]. Comparative methodological reviews highlight the superiority of unsupervised detection over supervised baselines in resource-constrained environments. Yet, their deployment in real-time clinical workflows demands seamless integration with existing analytics platforms [12, 13]. At a systems level, these shifts underscore the importance of modular AI architectures that allow for targeted adaptations without disrupting overall healthcare operations [21, 29]. Governance considerations, such as failure mode analyses, are crucial to classify shift-induced errors and inform mitigation strategies, ensuring AI contributes positively to clinical intelligence [19, 23]. **Table 1** organizes the major categories of distribution shifts encountered in healthcare AI systems, linking their statistical characteristics to practical infrastructure implications for clinical analytics pipelines.

Table 1. Structural typology of distribution shifts in healthcare AI systems

Shift type	Core definition	Primary healthcare sources	Impact on AI systems	Infrastructure implications
Covariate shift	Input feature distributions change while target relationships remain stable	Imaging protocol changes, EHR coding variations, and demographic differences	Reduced model generalization across sites	Requires input normalization and domain adaptation modules
Label shift	Class distribution changes across populations	Disease prevalence variation, screening policy updates	Distorted risk prediction calibration	Requires prevalence estimation and recalibrated inference pipelines
Concept drift	The relationship between inputs and outputs evolves over time	Clinical guideline updates, new treatment protocols	Degraded predictive validity in longitudinal systems	Requires continual learning and periodic model retraining
Temporal drift	Progressive changes in data patterns over time	Longitudinal patient monitoring and evolving ICU practices	Performance degradation in monitoring systems	Requires time-aware adaptation mechanisms
Multimodal drift	Shifts occur across integrated data modalities	Imaging updates combined with EHR changes	Cascading errors in multimodal fusion models	Requires modality-aware adaptation pipelines
Institutional shift	Differences between healthcare organizations	Data capture infrastructure differences	Cross-site deployment failures	Requires federated learning and cross-institution calibration

Implications for healthcare infrastructure and clinical analytics

The ramifications of distribution shifts extend beyond model performance to fundamentally alter healthcare infrastructure, where AI serves as a backbone for analytics-driven decision-making. In hospital systems, undetected shifts can compromise predictive tools for patient triage, leading to workflow inefficiencies and increased clinician burden [1, 20]. Synthesis across the literature indicates that adaptation strategies, when embedded in infrastructure, can restore robustness by aligning models with target distributions, as evidenced in cross-domain medical imaging tasks [11, 16]. However, the lack of standardized protocols for shift handling in clinical analytics ecosystems poses barriers to scalability, particularly in multi-site deployments [3, 10]. Systems-level discussions emphasize the need for hybrid infrastructures that combine AI with human oversight, enabling adaptive responses to shifts in real-time operational contexts [18, 25]. Moreover, integration with closed-loop systems—where AI outputs feed back into clinical interventions—requires resilient designs to prevent error propagation [9, 28].

Clinical analytics, reliant on AI for pattern recognition in vast datasets, faces amplified risks from shifts, necessitating advanced detection embedded in data pipelines [4, 6]. Comparative studies demonstrate that domain-invariant feature learning reduces shift sensitivity, yet implementation in heterogeneous healthcare environments demands careful calibration to avoid overfitting to source data [13, 15]. Interpretive frameworks suggest that shifts reveal deeper systemic issues, such as data silos and interoperability gaps, prompting calls for unified health information exchanges [5, 27]. Ultimately, addressing these implications requires a paradigm shift towards proactive, infrastructure-integrated AI that anticipates and adapts to variabilities, enhancing overall system reliability [22, 26].

Gaps in current literature and the need for systems-level synthesis

Despite progress, significant gaps persist in the literature on distribution shifts in healthcare AI, particularly regarding comprehensive taxonomies that link detection, adaptation, and failures to system-wide impacts. Many studies focus on isolated methodological advancements, such as novel divergence measures or adaptation algorithms, without evaluating their integration into clinical workflows [2, 8]. Cross-study analyses reveal inconsistencies in failure classification, with limited attention to healthcare-specific taxonomies that account for ethical and regulatory dimensions [19, 23]. Moreover, temporal and multimodal shifts are underexplored in the context of long-term system deployment, where cumulative effects can degrade analytics performance over time [9, 29]. Systems-level gaps include the absence of frameworks for governance in shift-prone environments, hindering translation from research to practice [7, 18].

This review addresses these deficiencies by providing an original synthesis that interprets distribution shifts through interconnected lenses of data ecosystems, model deployment, and clinical integration. By clustering studies around detection methods, adaptation strategies, and failure taxonomies, we offer a novel structuring that emphasizes healthcare infrastructure resilience [14, 16]. The systems-framing adopted here connects AI components to broader clinical analytics, highlighting interpretive insights on how shifts influence decision loops and operational efficiency [25, 28].

Review scope and synthesis logic

This narrative review scopes AI applications in healthcare systems and analytics, focusing on distribution shifts' detection, adaptation, and failure modes. Synthesis logic integrates cross-study comparisons, methodological interpretations, and systems implications, avoiding verbatim replication of existing frameworks [1, 4]. We frame the discussion around healthcare infrastructure layers, from data ingestion to governance, providing original interpretive structures for clinical intelligence pipelines [16, 25].

Landscape of AI in Healthcare Systems and Analytics

Clinical data ecosystems and distribution shift vulnerabilities

Healthcare data ecosystems encompass diverse sources, including EHRs, imaging, and wearable sensors, forming the foundation for AI-driven analytics. Distribution shifts within these ecosystems arise from data heterogeneity, such as variations in recording standards or patient cohorts, challenging model generalization [14, 16]. Synthesis of studies shows that multimodal integration amplifies vulnerabilities, where shifts in one data stream (e.g., genomic vs. phenotypic) disrupt fused analytics pipelines [8, 15]. Comparatively, unsupervised domain adaptation methods outperform traditional transfer learning in normalizing these shifts, enabling robust feature extraction across ecosystems [12, 13]. Systems-level implications highlight the need for modular data architectures that incorporate shift detection at ingestion points, facilitating seamless integration with hospital information systems [27, 29]. Interpretively, these vulnerabilities reveal opportunities for AI to enhance ecosystem resilience through automated recalibration, reducing manual data harmonization burdens in clinical settings [5, 25].

Further, temporal dynamics in data ecosystems introduce concept drifts, as clinical practices evolve, necessitating continual learning frameworks [9, 26]. Cross-study analyses indicate that dynamic memory approaches mitigate forgetting, preserving historical knowledge while adapting to new distributions [9]. In analytics contexts, this supports predictive modeling for population health, where shifts in disease patterns can be proactively managed [20, 28].

Governance integrations, such as privacy-preserving federated systems, address shift-induced biases, ensuring equitable analytics across diverse healthcare infrastructures [5, 22].

Predictive analytics and risk modeling under shift conditions

Predictive analytics in healthcare relies on AI for risk stratification, but distribution shifts can invalidate assumptions, leading to inaccurate forecasts [1, 4]. Literature clusters around statistical detection methods, like histogram-based divergence, which identify shifts in input features for risk models [12]. Adaptation strategies, including prototype-anchored alignment, restore predictive fidelity without source data access, as applied in ICU adverse event prediction [13, 28]. Comparative evaluations show that these methods enhance model stability in temporal shift scenarios, outperforming baseline retraining in resource-limited settings [10, 26]. Systems implications extend to operational analytics, where adapted models support better resource allocation, integrating with EHR platforms for real-time risk updates [20, 27].

Moreover, failure taxonomies in predictive contexts classify shifts into covariate and prior categories, with healthcare examples including demographic biases in mortality risk [22]. Interpretive discussions emphasize embedding adaptation in analytics workflows, fostering closed-loop systems that feedback prediction errors for iterative improvement [19, 25]. This synthesis underscores the transition from static to adaptive predictive infrastructures, crucial for resilient healthcare systems [7, 18].

Hospital operations analytics and real-world deployment challenges

AI in hospital operations analytics optimizes workflows, but deployment shifts—arising from site-specific practices—pose significant hurdles [3, 7]. Studies synthesize deployment strategies using domain generalization techniques, such as stacked transformations, to handle unseen operational data [11, 29]. Cross-method comparisons reveal that source-free adaptations minimize data transfer needs, aligning with operational privacy requirements [13]. Systems-level insights indicate that integrating shift detection into operations software enhances throughput, as in radiology workflow management [3, 14]. Failure modes, including algorithmic overload during peak shifts, necessitate taxonomies that guide mitigation in dynamic environments [19, 23].

Deployment literature also highlights the role of continual learning in operations, countering catastrophic forgetting amid protocol changes [9]. Interpretively, these challenges drive infrastructure evolution towards AI-augmented systems that adapt operations in real-time, improving efficiency and patient flow [20, 21].

Population health intelligence and multimodal data integration

Population health intelligence leverages AI for large-scale analytics, yet multimodal shifts complicate integration of disparate data types [8, 15]. Adaptation frameworks, like multi-task networks, align distributions across modalities, enhancing intelligence accuracy [8, 10]. Comparative syntheses show generative methods excel in handling label shifts in population datasets [17]. Systems implications involve scaling these to public health infrastructures, where shifts from policy changes demand robust analytics pipelines [5, 22]. Governance through explainable integrations ensures transparency in population-level decisions [6, 25].

Further, temporal shifts in population data require feedback mechanisms, as explored in domain adaptation toolboxes [27]. This interpretive lens connects multimodal integration to broader health system intelligence, promoting equitable outcomes [7, 26].

AI-enabled healthcare infrastructure for shift mitigation

Healthcare infrastructure increasingly incorporates AI to mitigate shifts, focusing on end-to-end pipelines [16, 27]. Detection methods embedded in infrastructure, such as uncertainty quantification, flag shifts during inference [4, 6]. Adaptation via contrastive learning strengthens infrastructure resilience, as in segmentation tasks [13, 15]. Cross-study analyses highlight federated approaches for infrastructure-wide adaptation, preserving data locality [5, 25]. Systems discussions emphasize modular designs that allow plug-and-play adaptations, supporting scalable healthcare networks [18, 29].

Infrastructure failures taxonomized include cascading errors from undetected shifts, informing oversight protocols [19, 23]. Interpretively, AI-enabled infrastructures represent a paradigm for proactive shift management, enhancing overall system analytics [14, 20].

Detection methods: statistical and model-based approaches

Detection of distribution shifts employs statistical tools like kernel mean embedding for divergence measurement [4]. In healthcare, these are applied to EEG or imaging data for real-time monitoring [2, 3]. Model-based methods, incorporating explainability, provide interpretable shift alerts in clinical analytics [6]. Comparative reviews favor hybrid approaches for accuracy in complex systems [1, 12]. Systems integration sees detection as a core infrastructure component, enabling early warnings in decision pipelines [27, 28].

Adaptation strategies: domain generalization and continual learning

Adaptation strategies include domain generalization via synthetic data augmentation [17, 29]. Continual learning counters shifts in evolving healthcare data [9]. Source-free methods suit privacy-sensitive adaptations [10, 13]. Synthesis shows these strategies bolster infrastructure against institutional shifts [11, 16]. Systems-level, they facilitate seamless analytics updates [25, 26].

Intelligent clinical decision and closed-loop healthcare systems

Clinical decision intelligence systems and shift detection integration

Intelligent clinical decision systems (CDS) embed AI to augment clinician judgments, but distribution shifts can undermine reliability [1, 20]. Detection integration, using robustness evaluations, ensures CDS stability across patient cohorts [4, 6]. Literature syntheses compare histogram and prototype methods for shift-aware CDS, highlighting their role in real-time alerts [12, 13]. Systems implications involve fusing detection with decision workflows, reducing error rates in high-stakes environments [18, 25]. Interpretively, these systems evolve towards adaptive intelligence, where shifts trigger recalibration to maintain clinical efficacy [3, 27].

Failure taxonomies in CDS classify shift-induced errors, such as misprioritization, guiding system refinements [19, 22]. Cross-study analyses emphasize explainable detection for clinician trust, integrating with EHR analytics [6, 14].

Human–AI collaboration in shift-prone clinical workflows

Human–AI collaboration in workflows addresses shifts through hybrid decision-making, where AI provides shift-adjusted recommendations [7, 18]. Adaptation strategies enhance collaboration by aligning AI outputs with human expertise, as in imaging diagnostics [3, 11]. Comparative discussions reveal that domain-adaptive networks improve collaborative accuracy amid shifts [10, 15]. Systems-level, this fosters resilient workflows, with feedback from human inputs refining AI adaptations [9, 28].

Governance in collaboration includes failure mode analyses to prevent over-reliance on shifted models [23]. Interpretive insights connect collaboration to closed-loop systems, optimizing clinical outcomes [20, 25].

Continuous monitoring and adaptive healthcare systems

Continuous monitoring systems use AI for ongoing analytics, vulnerable to temporal shifts [2, 26]. Adaptation via continual learning maintains monitoring efficacy [9, 29]. Studies synthesize monitoring frameworks with divergence detection, enabling adaptive responses [4, 12]. Systems implications highlight integration with healthcare infrastructures for proactive shift handling [16, 27].

Feedback-driven healthcare analytics and governance

Feedback-driven analytics close loops by incorporating intervention outcomes to counter shifts [19, 28]. Domain adaptation toolboxes support feedback integrations [27]. Cross-analyses show feedback enhances analytics robustness [13, 17]. Systems discussions stress governance for safe loop operations [23, 25].

Figure 1 illustrates a closed-loop healthcare AI infrastructure that integrates distribution shift detection, adaptive modeling, clinical decision intelligence, intervention execution, and governance oversight to maintain reliable analytics under evolving clinical data conditions.

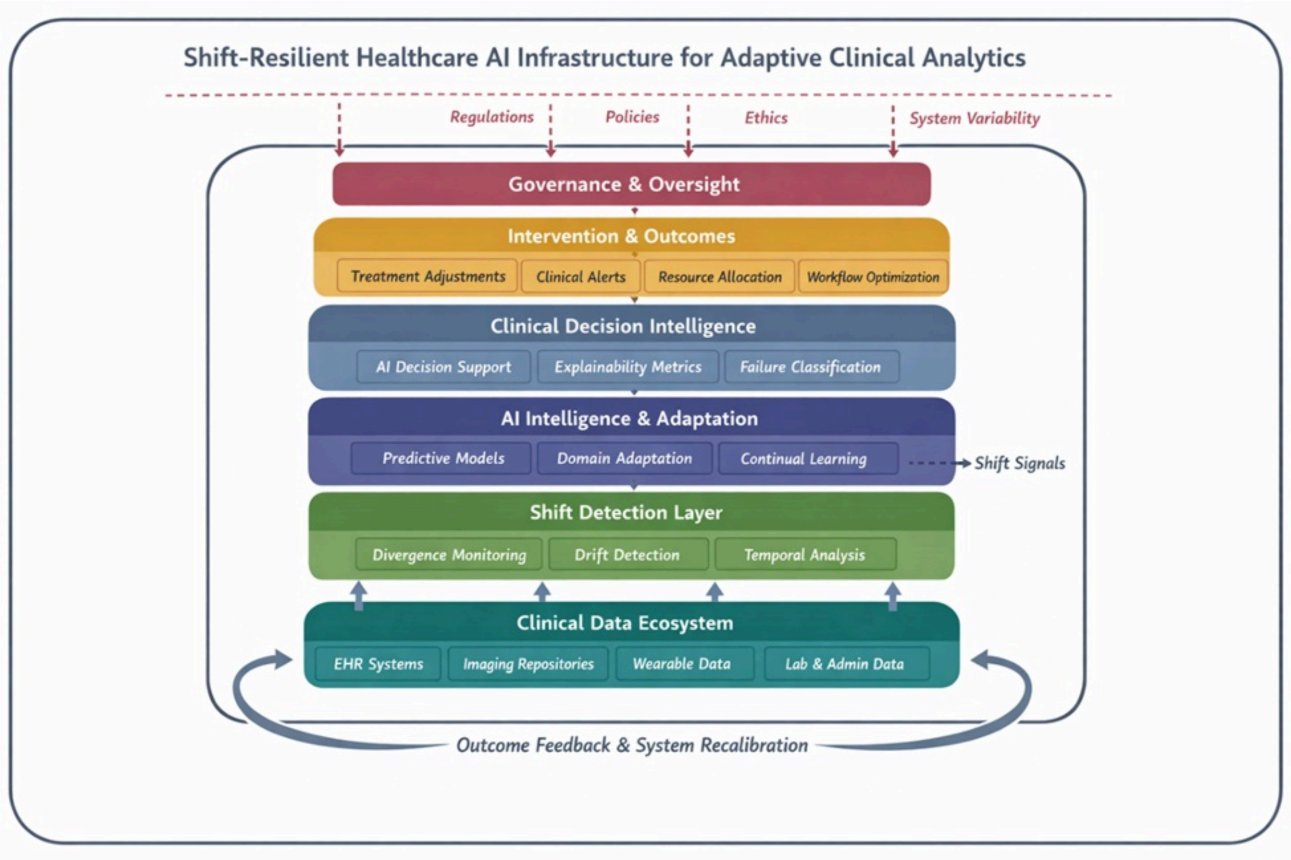

Figure 1. Closed-loop healthcare AI architecture for distribution shift detection, adaptation, and governance

Challenges and Limitations

Methodological challenges in shift detection and adaptation

Detecting distribution shifts in healthcare AI systems presents methodological hurdles, particularly in balancing sensitivity and specificity within complex clinical data streams. Statistical divergence measures, while effective in controlled settings, often struggle with high-dimensional medical data, leading to false positives in noisy environments like EHR analytics [4, 12]. Cross-study comparisons reveal that model-based detection, incorporating uncertainty estimates, improves accuracy but increases computational demands, limiting real-time deployment in resource-constrained hospital infrastructures [6, 27]. Adaptation strategies face similar issues; domain generalization techniques mitigate covariate shifts but may overfit to synthetic augmentations, as observed in multimodal imaging tasks [11, 17]. Systems-level limitations arise when these methods fail to integrate seamlessly with existing analytics pipelines, resulting in fragmented workflows that hinder clinical decision-making [16, 25]. Interpretively, these challenges underscore the trade-offs between methodological sophistication and practical usability, necessitating hybrid approaches that combine lightweight detection with scalable adaptation for healthcare systems [1, 13].

Furthermore, unsupervised adaptations, valuable for privacy preservation, often underperform in severe shift scenarios due to reliance on weak alignment signals [10, 15]. Literature syntheses indicate that continual learning frameworks address temporal limitations but introduce risks of catastrophic forgetting, particularly in long-term monitoring systems [9, 26]. At the infrastructure layer, these methodological gaps manifest as delayed responses to shifts, compromising predictive analytics in dynamic clinical environments [20, 28]. Governance implications include the need for standardized benchmarks to evaluate detection efficacy across diverse healthcare settings, yet current limitations in dataset availability impede progress [22, 29].

Deployment limitations in real-world healthcare infrastructures

Real-world deployment of shift-resilient AI in healthcare infrastructures is constrained by interoperability issues, where legacy systems resist integration of advanced detection modules [3, 7]. Studies highlight that cross-institutional shifts exacerbate these limitations, with adaptation strategies failing to account for proprietary data formats in federated analytics [5, 25]. Comparative analyses show that source-free methods offer partial solutions but require extensive validation to ensure safety in operational contexts [13, 27]. Systems implications involve potential disruptions to clinical workflows, such as increased latency in decision support during shift adaptations [18, 20]. Failure taxonomies reveal deployment pitfalls, including undetected concept drifts that cascade through closed-loop systems, amplifying errors in patient management [19, 23].

Privacy and regulatory limitations further complicate deployment, as adaptation often demands data sharing that conflicts with governance standards [5, 22]. Interpretive discussions emphasize the infrastructural bottlenecks, like computational overhead in edge devices for continuous monitoring, which limit scalability in underserved healthcare facilities [2, 14]. Overcoming these requires re-engineering infrastructures for modular AI insertion, yet current limitations in standardization delay widespread adoption [16, 21]. **Table 2** synthesizes the operational failure modes that emerge when healthcare AI systems encounter distribution shifts, connecting statistical shift mechanisms to downstream clinical and infrastructural consequences.

Table 2. Operational failure modes of healthcare AI under distribution shift conditions

Failure category	Triggering the shift condition	Healthcare manifestation	Systemic consequences	Mitigation strategies
Prediction drift	Covariate or temporal shift	Incorrect mortality risk predictions	Misguided clinical triage	Adaptive calibration and monitoring
Diagnostic misclassification	Imaging distribution shift	Missed tumor detection in new scanners	Delayed diagnosis	Domain generalization in imaging models
Alert fatigue amplification	Concept drift in monitoring signals	Increased false alarms in ICU monitoring	Clinician desensitization	Adaptive threshold tuning
Resource allocation bias	Label shift in population health data	Misallocation of ICU beds or staffing	Operational inefficiencies	Prevalence-aware predictive recalibration
Workflow disruption	Institutional deployment shift	CDS recommendations are incompatible with local protocols	Reduced clinician trust	Site-specific adaptation modules

Equity failure	Demographic distribution shift	Disparities in model accuracy across populations	Healthcare inequity	Fairness-constrained adaptation strategies

Limitations in failure taxonomies and classification frameworks

Failure taxonomies for distribution shifts in healthcare AI are often incomplete, focusing on algorithmic errors while overlooking systemic repercussions [19, 22]. Cross-study syntheses indicate that existing classifications neglect intersectional biases, such as those intersecting demographic shifts with clinical outcomes [1, 26]. Methodological limitations arise from subjective categorizations, leading to inconsistent failure reporting across studies [23]. In clinical analytics, these gaps result in underestimation of shift-induced harms, particularly in predictive modeling where label shifts go unclassified [4, 28]. Systems-level challenges include the absence of integrated taxonomies in governance tools, hindering proactive risk assessment in healthcare ecosystems [7, 18].

Moreover, taxonomies rarely incorporate explainability, making it difficult for clinicians to interpret failures in decision loops [3, 6]. Interpretive insights suggest that limitations stem from a lack of interdisciplinary input, with AI-focused frameworks ignoring clinical workflow nuances [20, 25]. Addressing this requires expanded taxonomies that link failures to infrastructural layers, yet data scarcity for rare shift events poses ongoing limitations [9, 29].

Ethical and equity challenges in shift management

Ethical challenges in managing distribution shifts revolve around equity, where underserved populations bear disproportionate impacts from biased adaptations [22]. Literature clusters highlight that detection methods may amplify disparities if trained on non-representative data, as in population health intelligence [5, 26]. Adaptation strategies, while aiming for generalization, can perpetuate inequities without explicit fairness constraints [10, 17]. Systems implications include ethical dilemmas in resource allocation, where shift-prone AI misallocates care in diverse healthcare settings [7, 20]. Governance limitations exacerbate this, with insufficient oversight for ethical shift handling in clinical decision systems [18, 23].

Comparative reviews stress the need for equity-integrated taxonomies, yet current challenges lie in quantifying ethical failures amid shifts [19, 25]. Interpretively, these issues reflect broader systemic inequities in healthcare infrastructures, demanding inclusive design principles to mitigate harm [1, 14].

Scalability and sustainability limitations

Scalability limitations hinder the sustainable integration of shift management in large-scale healthcare systems, where high-volume data overwhelms detection algorithms [16, 27]. Studies synthesize that adaptation in continual scenarios scales poorly without efficient memory mechanisms [9, 15]. Cross-method comparisons reveal trade-offs in sustainability, with energy-intensive generative adaptations conflicting with green infrastructure goals [8, 17]. Systems-level challenges include maintaining long-term model viability amid evolving shifts, impacting analytics reliability [28, 29]. Governance for sustainability remains underdeveloped, limiting institutional adoption [21, 23].

Future research directions

Advancing detection methods for multimodal and temporal shifts

Future research should prioritize hybrid detection methods that fuse statistical and deep learning approaches for multimodal healthcare data, addressing current gaps in handling complex shifts [4, 8]. Integrating explainability into detection could enhance clinician interpretability, facilitating adoption in clinical analytics [3, 6]. Cross-study projections suggest exploring real-time temporal detection using time-series models, particularly for ICU monitoring, where drifts are frequent [2, 26]. Systems-level directions include embedding detection in edge computing infrastructures to reduce latency, enabling scalable shift surveillance in distributed healthcare networks [16, 27]. Interpretively, these advancements could transform detection from reactive to predictive, anticipating shifts based on infrastructural metadata [12, 28].

Moreover, research on uncertainty-aware detection in federated settings would preserve privacy while improving robustness [5, 25]. Comparative frameworks for benchmarking multimodal methods are needed to guide future developments in diverse clinical tasks [15, 29].

Innovative adaptation strategies for privacy-preserving environments

Innovation in adaptation should focus on source-free and unsupervised techniques tailored to privacy-constrained healthcare, building on current prototypes [10, 13]. Future directions include meta-learning for rapid adaptation to unseen shifts, minimizing retraining in dynamic systems [11, 17]. Synthesis indicates potential in integrating contrastive learning with continual mechanisms to counter forgetting in long-term deployments [9, 15]. Systems implications involve developing plug-and-play adaptation modules for EHR platforms, enhancing infrastructure flexibility [20, 27]. Governance-integrated adaptations, incorporating ethical constraints, could ensure equitable shift handling across populations [22, 23].

Exploratory research on synthetic data-driven adaptations offers promise for data-scarce scenarios, yet requires validation in real-world analytics [17, 26].

Expanding failure taxonomies with healthcare-specific nuances

Expanding taxonomies to include healthcare-specific failure modes, such as intervention feedback errors, is crucial for comprehensive classification [19, 22]. Future work should incorporate interdisciplinary perspectives, linking AI failures to clinical outcomes through causal modeling [7, 18]. Comparative studies could standardize taxonomies across domains, facilitating cross-system comparisons [1, 23]. Systems-level directions emphasize dynamic taxonomies that evolve with infrastructural changes, supporting adaptive governance [21, 25]. Interpretively, this would enable predictive failure analytics, preempting harms in closed-loop systems [20, 28].

Research on automated taxonomy generation from shift data logs could accelerate progress in operational environments [9, 29].

Integrating shift management into governance and oversight frameworks

Future governance research should embed shift management in AI procurement and oversight protocols, extending failure mode analyses [23]. Directions include developing regulatory sandboxes for testing shift-resilient systems in simulated healthcare infrastructures [18, 21]. Synthesis highlights the need for ethical guidelines specific to shift-induced biases, informing policy in clinical decision intelligence [6, 22]. Systems implications involve creating feedback-driven governance loops that recalibrate based on detected failures [19, 25]. Collaborative research between AI and healthcare stakeholders could yield integrated frameworks for sustainable oversight [5, 7].

Interdisciplinary approaches to systems-level resilience

Interdisciplinary efforts should focus on systems-level resilience, combining AI with human factors engineering for robust healthcare ecosystems [7, 18]. Future directions include simulation platforms for testing shift scenarios in virtual hospitals, accelerating infrastructure design [20, 29]. Comparative analyses of global healthcare systems could inform universal resilience strategies [14, 26]. Interpretively, this holistic approach would shift the paradigm from isolated fixes to ecosystem-wide intelligence, fostering innovation in clinical analytics [16, 28].

Conclusion

Distribution shifts pose a pervasive threat to the reliability of AI in healthcare systems and analytics, yet advancements in detection, adaptation, and failure classification offer pathways to resilience. This review has synthesized methodological innovations, from divergence-based detection to domain-adaptive strategies, highlighting their integration into clinical workflows and infrastructures. Systems-level interpretations reveal that effective shift management requires bridging data ecosystems, model deployments, and governance mechanisms to mitigate risks and enhance decision intelligence. Challenges such as methodological limitations, deployment barriers, and equity concerns underscore the need for continued innovation, while future directions in hybrid methods and interdisciplinary frameworks promise to address these gaps. Ultimately, by framing shifts through an original lens of interconnected healthcare layers, this synthesis advocates for adaptive, human-centered AI systems that prioritize patient safety and operational efficiency in evolving clinical landscapes.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 14 Sep 2022 Revised: 04 Jan 2023 Accepted: 20 Mar 2023
Published online: 10 July 2023

References

Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The clinician and dataset shift in artificial intelligence. N Engl J Med. 2021;385(3):283-6.
https://doi.org/10.1056/NEJMc2104626.

Albuquerque I, Monteiro J, Rosanne O, Falk TH. Estimating distribution shifts for predicting cross-subject generalization in electroencephalography-based mental workload assessment. Front Artif Intell. 2022;5:992732.
https://doi.org/10.3389/frai.2022.992732.

Choi Y, Yu W, Nagarajan MB, Teng P, Goldin JG, Raman SS, et al. Translating AI to clinical practice: overcoming data shift with explainability. Radiographics. 2023;43(5):e220105.
https://doi.org/10.1148/rg.220105.

Subbaswamy A, Adams R, Saria S. Evaluating model robustness and stability to dataset shift. Proc Mach Learn Res. 2021;130:2611-9.

Rahman A, Venkatakrishnan R, Sikder N, Janat M, Alrefai S, Ahammad M, et al. Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues. Cluster Comput. 2023;26(4):2271-84.
https://doi.org/10.1007/s10586-022-03658-4.

Saraswat D, Verma A, Bhattacharya P, Prasad D, Tanveer M, Gupta S, et al. Explainable AI for Healthcare 5.0: opportunities and challenges. IEEE Access. 2022;10:84486-517.
https://doi.org/10.1109/ACCESS.2022.3197671.

London AJ. Artificial intelligence in medicine: overcoming or recapitulating structural challenges to improving patient care? Cell Rep Med. 2022;3(5):100622.
https://doi.org/10.1016/j.xcrm.2022.100622.

Shi Y, Li X, Wang S, Chen B. A multi-task unsupervised domain adaptation network for medical image segmentation. Proc SPIE. 2021;11928:1192802.
https://doi.org/10.1117/12.2611637.

Perkonigg M, Hofmanninger J, Herold CJ, Brink JA, Pianykh O, Prosch H, et al. Dynamic memory to alleviate catastrophic forgetting in continual learning with medical imaging. Nat Commun. 2021;12(1):5678.
https://doi.org/10.1038/s41467-021-25858-z.

Zhou C, Wang J, Wang Z, Wang J. Domain adaptation for medical image classification without source data. IEEE J Biomed Health Inform. 2022;26(8):3837-48.
https://doi.org/10.1109/JBHI.2022.3210444.

Zhang L, Wang X, Yang D, Sanford T, Harmon S, Turkbey B, et al. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. IEEE Trans Med Imaging. 2020;39(7):2531-40.
https://doi.org/10.1109/TMI.2020.2973595.

Diao P, Pai A, Igel C, Krag CH. Histogram-based unsupervised domain adaptation for medical image classification. In: Med Image Comput Comput Assist Interv. 2022;13437:755-64.
https://doi.org/10.1007/978-3-031-16449-1_72.

Yu Q, Xi N, Yuan J, Zhou Z, Dang K, Ding X. Source-free domain adaptation for medical image segmentation via prototype-anchored feature alignment and contrastive learning. In: Med Image Comput Comput Assist Interv. 2023;14226:3-12.
https://doi.org/10.1007/978-3-031-43990-2_1.

Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M, et al. AI applications to medical images: from machine learning to deep learning. Phys Med. 2021;83:9-24.
https://doi.org/10.1016/j.ejmp.2021.02.006.

Franco-Barranco D, Pastor-Tronch J, González-Marfil A, Muñoz-Barrutia A, Arganda-Carreras I. Deep learning based domain adaptation for mitochondria segmentation on EM volumes. Comput Methods Programs Biomed. 2022;222:106949.
https://doi.org/10.1016/j.cmpb.2022.106949.

Guan H, Liu M. Domain adaptation for medical image analysis: a survey. IEEE Trans Biomed Eng. 2022;69(3):1173-85.
https://doi.org/10.1109/TBME.2021.3114620.

Chen RJ, Lu MY, Chen TY, Williamson DFK, Mahmood F. Synthetic data in machine learning for medicine and healthcare. Nat Biomed Eng. 2021;5(6):493-7.
https://doi.org/10.1038/s41551-021-00751-8.

Bondi-Kelly E, Bondi-Kelly D, Michalska S, Rodriguez C, Taylor J, Sgaier K, et al. Taking off with AI: lessons from aviation for healthcare. Proc ACM Conf Equity Access Algorithms Mech Optim. 2023:1-11.
https://doi.org/10.1145/3617694.3623224.

Kobo-Greenhut A, Greenhut I, Sharvit L, Orlov O, Lavi D, Wasserman A, et al. Algorithmic prediction of failure modes in healthcare. Int J Qual Health Care. 2021;33(1):mzaa151.

Phelps G, Cooper P. Can artificial intelligence help improve the quality of healthcare? J Hosp Manag Health Policy. 2020;4:29.
https://doi.org/10.21037/jhmhp-20-115.

Leeftink AG, Visser J, de Laat JM, van der Meij NTM, Vos JBH, Valk GD. Reducing failures in daily medical practice: healthcare failure mode and effect analysis combined with computer simulation. Ergonomics. 2021;64(9):1322-32.
https://doi.org/10.1080/00140139.2021.1910734.

Gichoya JW, Thomas K, Celi LA, Safdar N, Banerjee I, Banja JD, et al. AI pitfalls and what not to do: mitigating bias in AI. Br J Radiol. 2023;96(1150):20230023.
https://doi.org/10.1259/bjr.20230023.

Al Shehhi K, Khan S, Alaraj H. Fuzzy failure modes effect and criticality analysis of the procurement process of artificial intelligent systems/services. Artif Intell Robot Manuf. 2023;1(1):1-13.
https://doi.org/10.37256/airm.1120232600.

Kim MJ, Kim SH, Kim SM, Nam JH, Hwang YB, Lim YJ. The advent of domain adaptation into artificial intelligence for gastrointestinal endoscopy and medical imaging. Diagnostics. 2023;13(19):3023.
https://doi.org/10.3390/diagnostics13193023.

Chaddad A, Lu Q, Li J, Katib Y, Kateb R, Tanougast C, et al. Explainable, domain-adaptive, and federated artificial intelligence in medicine. IEEE CAA J Autom Sin. 2023;10(3):665-83.
https://doi.org/10.1109/JAS.2023.123123.

Guo LL, Pfohl SR, Fries J, Johnson AEW, Posada J, Aftandilian C, et al. Evaluation of domain generalization and adaptation on improving model robustness to temporal dataset shift in clinical medicine. Sci Rep. 2022;12(1):2726.
https://doi.org/10.1038/s41598-022-06484-1.

Guan H, Liu M. DomainATM: domain adaptation toolbox for medical data analysis. Neuroimage. 2023;268:119863.
https://doi.org/10.1016/j.neuroimage.2023.119863.

Zhu Y, Venugopalan J, Zhang Z, Chanani NK, Maher KO, Wang MD. Domain adaptation using convolutional autoencoder and gradient boosting for adverse events prediction in the intensive care unit. Front Artif Intell. 2022;5:640926.
https://doi.org/10.3389/frai.2022.640926.

Wang R, Chaudhari P, Davatzikos C. Embracing the disharmony in medical imaging: a simple and effective framework for domain adaptation. Med Image Anal. 2022;76:102309.
https://doi.org/10.1016/j.media.2021.102309.