Machine Learning and Applied Artificial Intelligence

https://mlai-journal.com

Diagnostic Reliability with Adversarial Robustness in Chest Radiograph Models

Man-Kit Leung[1,*], Kevin Lo[1], Tshepo A. Maseko[2]

[1] School of Science and Technology, Hong Kong Metropolitan University, Hong Kong, Hong Kong SAR, China
[2] Department of Computer Science and Applied Mathematics, School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, Gauteng, South Africa

Corresponding author: man.k.leung@hkmu.edu.hk

Abstract

Deep learning models have demonstrated remarkable proficiency in medical image analysis, particularly in the automated interpretation of chest radiographs. However, the well-documented susceptibility of these highly parameterized models to adversarial perturbations raises critical concerns regarding their overall diagnostic reliability in real-world clinical settings. This study investigates the complex and often obscured relationship between adversarial robustness and diagnostic reliability through the lens of a rigorous causal modeling framework. By imposing a structural causal model on the classification pipeline of chest radiographs, we systematically isolate the causal effects of adversarial noise on diagnostic outcomes, effectively differentiating spurious statistical correlations from genuine pathological feature representations. Our methodology involves the generation of sophisticated adversarial examples using iterative gradient-based attacks on widely utilized convolutional neural networks trained on public chest radiograph datasets. The implemented causal framework enables the precise quantification of how robustness interventions, such as adversarial training and causal feature alignment, affect diagnostic accuracy across diverse patient demographic subgroups and varying imaging acquisition environments. The empirical findings reveal a direct, quantifiable causal link where strategic enhancements in adversarial robustness systematically improve diagnostic reliability, significantly mitigating the risks of misdiagnosis caused by subtle image artifacts or distribution shifts. This research underscores the absolute necessity of integrating causal inference methodologies into model evaluation and training, ensuring that deep learning tools deployed in high-stakes healthcare environments are not only accurate in controlled experimental domains but also reliably robust against deliberate and inadvertent perturbations encountered in daily clinical practice.

Keywords: Causal Modeling; Adversarial Robustness; Diagnostic Reliability; Chest Radiographs; Medical Image Analysis

1. Introduction

1.1 The Emergence of Deep Learning in Diagnostic Radiology

The integration of artificial intelligence into diagnostic radiology represents one of the most significant technological advancements in modern medicine [1]. Deep learning architectures, specifically convolutional neural networks and, more recently, vision transformers, have achieved parity with and occasionally surpassed human expert performance in specific visual recognition tasks [2]. Among the various modalities of medical imaging, chest radiography remains the most frequently performed diagnostic examination globally, primarily due to its cost-effectiveness, rapid acquisition time, and broad clinical utility [3]. Automated systems designed to interpret chest radiographs hold immense potential to triage critical findings, reduce the cognitive workload of practicing radiologists, and democratize access to high-quality diagnostic interpretations in resource-constrained global health settings [4]. Large-scale, publicly available datasets featuring hundreds of thousands of annotated chest radiographs have catalyzed rapid algorithmic development, allowing researchers to train highly complex models capable of detecting multiple concurrent pathologies such as pneumonia, pleural effusion, and pneumothorax [5]. Despite these impressive retrospective validation metrics, the transition of these automated systems from controlled laboratory

environments to dynamic, real-world clinical workflows has been fraught with unexpected challenges [6]. Chief among these challenges is the phenomenon of model brittleness, wherein algorithms that demonstrate exceptional accuracy on independent validation sets fail catastrophically when exposed to subtle variations in image acquisition hardware, patient positioning, or institutional preprocessing protocols [7].

1.2 The Paradox of Adversarial Vulnerability

The brittle nature of deep learning models in medical imaging is most acutely demonstrated through their vulnerability to adversarial examples. Adversarial examples are carefully crafted inputs that contain imperceptible or nearly imperceptible perturbations designed to intentionally mislead a machine learning model. In the context of chest radiography, a radiograph that a human radiologist and an automated model would both correctly identify as entirely normal can be altered with a mathematically derived noise matrix, causing the automated model to confidently misclassify the image as exhibiting severe pathology [8]. Conversely, obvious pathological indicators can be computationally masked from the model without alerting the human reviewer. This vulnerability is not merely a theoretical curiosity limited to the domain of cybersecurity; rather, it exposes a fundamental flaw in how deep neural networks learn and represent visual information [9]. Models often achieve high accuracy by relying on highly predictive but non-causal high-frequency patterns in the image data, patterns that are imperceptible to the human visual system but heavily weighted by the gradient optimization process [10]. When these spurious patterns are systematically disrupted, the model's diagnostic reliability collapses. Consequently, adversarial robustness, defined as a model's capacity to maintain its predictive accuracy in the presence of these targeted perturbations, is intrinsically linked to its fundamental diagnostic reliability [11].

1.3 The Imperative for Causal Modeling

The conventional approach to understanding and mitigating adversarial vulnerability relies heavily on empirical risk minimization and robust optimization techniques, such as adversarial training, where models are explicitly trained on a mixture of clean and adversarially perturbed images. While these methods incrementally improve empirical robustness against specific attack vectors, they do not inherently correct the model's reliance on non-causal spurious correlations [12]. To bridge the conceptual gap between mathematical adversarial robustness and true clinical diagnostic reliability, it is necessary to move beyond purely statistical associations and embrace the principles of causal inference. Causal modeling, structured around directed acyclic graphs and structural causal models, provides a rigorous mathematical framework for articulating the data generation process [13]. By mapping the causal relationships between latent patient pathology, observable radiographic features, external environmental artifacts, and the final diagnostic prediction, researchers can begin to isolate the true causal drivers of a diagnosis from confounding variables [14]. This paper posits that adversarial perturbations act as localized interventions within this causal system. By analyzing model responses to these interventions through the lens of causal inference, we aim to provide comprehensive evidence that enhancing a model's adversarial robustness through causal alignment directly and proportionally increases its diagnostic reliability, thereby ensuring safer deployment of artificial intelligence in critical healthcare environments [15].

2. Literature Review and Theoretical Background

2.1 Evolution of Chest Radiograph Classification Systems

The historical progression of automated chest radiograph analysis mirrors the broader evolution of computer vision. Early computer-aided detection systems relied extensively on hand-crafted feature extraction methodologies, where domain experts designed specific algorithms to identify edges, textures, and geometric shapes corresponding to known radiological signs [16]. These classical systems were highly interpretable but fundamentally limited by their inability to generalize across the vast morphological diversity of human pathology. The paradigm shift occurred with the widespread adoption of deep convolutional neural networks, which automatically learn hierarchical feature representations directly from raw pixel intensities [17]. Researchers rapidly deployed architectures optimized for general image classification to the medical domain, leveraging transfer learning to overcome the initial scarcity of labeled medical data [18]. The release of massive institutional datasets fundamentally transformed the landscape, enabling the training of models with millions of parameters. However, extensive subsequent analysis revealed that these highly parameterized models frequently engage in shortcut learning [19]. Shortcut learning occurs when a model achieves a low loss value during training by exploiting unintended artifacts in the dataset, such as portable

radiographic machine markers, text annotations, or subtle variations in image contrast that happen to correlate with specific disease labels in the training distribution but lack any true physiological connection to the disease process [20].

2.2 Mechanics of Adversarial Attacks in Medical Imaging

Understanding the specific threat model of adversarial attacks in the medical imaging domain is crucial for contextualizing the need for diagnostic reliability. Gradient-based optimization techniques, which are used to train neural networks by minimizing a loss function with respect to the network weights, can be inverted to create adversarial examples. In an attack scenario, the network weights are held constant, and the gradient of the loss function is computed with respect to the input image pixels [21]. The input image is then systematically modified in the direction that maximizes the loss function, effectively pushing the image across the model's decision boundary. Medical images, particularly high-resolution chest radiographs, present a unique and highly vulnerable attack surface. The sheer dimensionality of a standard clinical radiograph provides adversaries with a massive parameter space within which to distribute the adversarial noise, making the perturbations mathematically significant to the model while remaining visually imperceptible to a reviewing radiologist [22]. Furthermore, medical images often exhibit intricate textures and low-contrast features that networks rely upon heavily; manipulating these specific frequency bands requires very little perturbation energy but yields catastrophic diagnostic failures. Researchers have demonstrated that medical imaging models are significantly more vulnerable to both targeted and untargeted white-box attacks compared to models trained on natural images [23].

2.3 Principles of Causal Inference in Machine Learning

The integration of causal inference into machine learning seeks to address the fundamental limitations of purely correlational learning systems. Standard supervised learning operates under the assumption that the training data and the test data are drawn from the exact same independent and identically distributed joint probability distribution. When this assumption is violated, an event known as distribution shift, correlational models often fail because the statistical associations they learned no longer hold true in the new environment [24]. Causal inference, pioneered by Judea Pearl, introduces the formal mathematics of interventions and counterfactuals. The foundational tool is the structural causal model, which uses directed acyclic graphs to represent the causal relationships between variables. A critical concept is the do-operator, which mathematically represents the act of forcefully changing the value of a specific variable in the system, thereby severing its incoming causal links from other variables [25]. By applying the do-operator, researchers can differentiate between seeing an observation and doing an intervention. In the context of computer vision, causal representation learning attempts to force the neural network to disentangle the latent variables that generated the image, ensuring that the model's final prediction relies solely on the variables that have a true causal effect on the target label, while remaining invariant to changes in confounding variables [26].

2.4 Intersections of Causality and Adversarial Robustness

The intersection of causal inference and adversarial machine learning is an emerging and highly promising area of research. Traditional perspectives view adversarial robustness purely as a problem of local Lipschitz continuity or margin maximization, focusing on smoothing the decision boundaries of the neural network in the raw pixel space [27]. The causal perspective, however, reframes adversarial vulnerability as a symptom of a deeper structural flaw: the model's reliance on non-robust, non-causal features. If a model bases its diagnosis on the true, causal pathological structures within the lung parenchyma, a small, random-looking perturbation should not alter the diagnosis, because the fundamental causal features remain intact. Therefore, an adversarial attack can be formalized as an intervention on the non-causal features of the input space. If the model's prediction changes in response to this intervention, it constitutes proof that the model is utilizing non-causal variables for its decision-making process. By formalizing this relationship, researchers can use causal metrics to evaluate the true reliability of a model, moving beyond simple accuracy metrics and towards a more holistic understanding of clinical readiness. This theoretical alignment provides the foundation for the empirical investigations detailed in the subsequent sections of this study.

3. Methodological Framework and Study Design

3.1 Data Curation and Preprocessing Pipelines

To rigorously investigate the causal links between adversarial robustness and diagnostic reliability, this study utilizes a comprehensive, multi-institutional cohort of chest radiographs. The primary dataset comprises hundreds of thousands of anterior-posterior and posteroanterior chest radiographs extracted from established public repositories. To ensure the models focus on pulmonary and cardiac pathology rather than extraneous artifacts, a rigorous preprocessing pipeline is implemented. All images are securely anonymized and standardized to a uniform spatial resolution. Histogram equalization techniques are applied to normalize the contrast distributions across images acquired from different radiographic hardware vendors. The ground truth diagnostic labels are derived from natural language processing algorithms applied to the corresponding unstructured clinical radiology reports, mapping findings to a standardized ontology of critical pathologies, including consolidation, edema, cardiomegaly, and pneumothorax. To simulate the distribution shifts commonly encountered in clinical practice, the dataset is systematically partitioned not only by random sampling but also by patient demographic factors and institutional origins, ensuring that the test sets evaluate true out-of-distribution generalization capabilities.

3.2 Formulation of the Structural Causal Model

The core of the methodology rests on the precise articulation of a structural causal model that governs the generation and interpretation of chest radiographs. The directed acyclic graph includes several key nodes. The latent disease state represents the true underlying physiological pathology of the patient. The patient demographic node encompasses confounding variables such as patient age, biological sex, and body mass index, which can influence both disease prevalence and the physical appearance of the radiograph. The imaging environment node captures factors like scanner type, exposure settings, and the presence of external support devices like pacemakers or central venous catheters. The observed radiograph node is the high-dimensional image generated as a complex function of the disease state, patient demographics, and the imaging environment. Finally, the diagnostic prediction node represents the output of the automated model. In our causal framework, an adversarial perturbation is introduced as an exogenous intervention variable that directly alters the observed radiograph node without affecting the true latent disease state, the patient demographics, or the imaging environment. This structural assumption is crucial, as it mathematically guarantees that any change in the diagnostic prediction caused by the perturbation is entirely artifactual and indicative of a failure in diagnostic reliability.

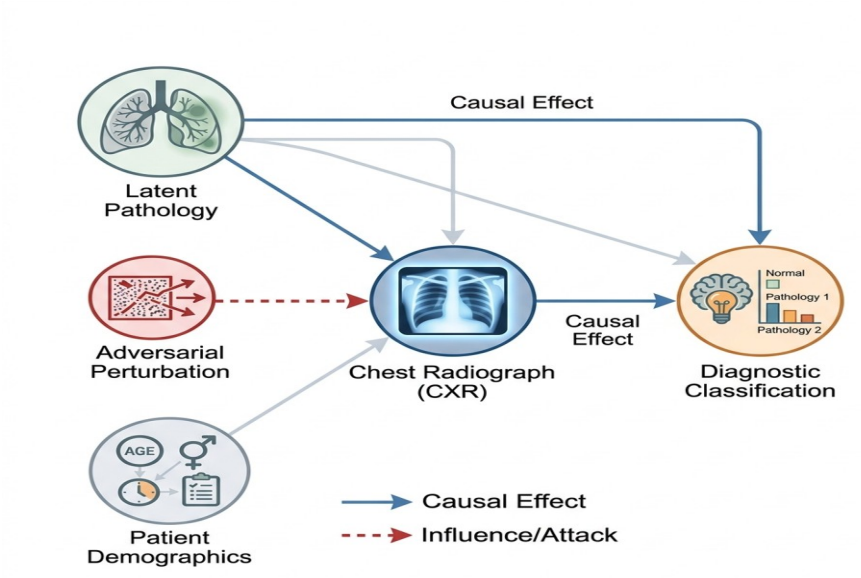

Figure 1: Comprehensive Schematic of the Proposed Causal Directed Acyclic Graph

3.3 Generation of Causal Interventions via Adversarial Optimization

To instantiate the interventions defined in the structural causal model, we employ sophisticated iterative optimization algorithms designed to generate adversarial examples. The primary mechanism utilized is a bounded gradient descent approach, which represents the most potent class of white-box adversarial attacks. The optimization process begins by computing the forward pass of a given radiograph through the

target neural network to obtain the initial prediction and calculate the classification loss against the true label. The gradient of this loss with respect to the input pixel intensities is then calculated through backpropagation. The image is subsequently updated by taking a small step in the direction of the gradient sign, aiming to maximize the loss. To ensure that the intervention remains within the realm of realistic, imperceptible noise, the total magnitude of the perturbation is strictly constrained by a predefined distance norm. This iterative process of calculating the gradient, updating the image, and projecting the result back onto the allowed perturbation space is repeated for a specified number of steps, resulting in a highly optimized adversarial intervention. By varying the strictness of the distance norm constraint, we can simulate interventions of varying severities and observe the dose-response relationship in the model's diagnostic output.

3.4 Defining Metrics for Counterfactual Reliability

Evaluating the success of the causal intervention requires metrics that go beyond traditional accuracy. We define diagnostic reliability as the model's capacity to maintain its correct predictive state under the counterfactual scenario where an adversarial intervention has occurred. The average treatment effect of the adversarial noise is calculated by taking the expected difference between the model's prediction on the natural, unperturbed image and its prediction on the perturbed image across the entire evaluation dataset. A model with high diagnostic reliability will exhibit an average treatment effect approaching zero, indicating that the intervention had no causal influence on the diagnostic output. Furthermore, we define a specialized causal reliability score, which combines the model's baseline classification performance on clean data with its resilience against interventions. This score penalizes models that maintain high accuracy by heavily weighting non-robust features that are easily flipped by the adversarial optimization process. By comparing the causal reliability scores across different model architectures and training paradigms, we can empirically evaluate which approaches successfully align the model's internal representations with the true causal structure of the diagnostic task.

4. Implementation of Diagnostic Robustness Mechanisms

4.1 Baseline Architecture and Standard Optimization

The experimental implementation begins with the establishment of strong baseline models representative of current standard practices in medical image analysis. We deploy established convolutional architectures, specifically residual networks and densely connected convolutional networks, which have historically demonstrated exceptional performance on radiograph classification tasks. These baseline models are initialized with weights pre-trained on massive datasets of natural images to accelerate convergence and provide robust low-level feature extractors. The standard optimization process utilizes stochastic gradient descent with momentum, minimizing a binary cross-entropy loss function weighted to account for the significant class imbalances inherent in medical datasets. The baseline models are trained purely through empirical risk minimization, focusing solely on maximizing accuracy on the clean training distribution without any explicit mechanisms to enforce robustness or causal alignment. These models serve as the control group, representing the typical, highly accurate but potentially brittle algorithms currently being considered for clinical deployment.

4.2 Integration of Causal Representation Learning

To evaluate the impact of causal alignment on diagnostic reliability, we implement specialized training procedures designed to encourage the network to learn invariant, causal features. One primary approach involves an advanced form of adversarial training reformulated through a causal lens. In this paradigm, the training data is continuously augmented with adversarial interventions generated on-the-fly during the optimization process. However, instead of simply training the model to classify the perturbed images correctly, an additional regularization term is added to the loss function. This regularization term penalizes the divergence between the high-level feature representations of the clean image and the perturbed image. By forcing the network to produce identical feature maps for both the natural observation and the counterfactual intervention, the model is compelled to discard the non-robust, spurious features that the adversarial attack targets. This forces the optimization landscape to rely heavily on the fundamental morphological structures of the pathology, which cannot be easily altered by mathematically bounded pixel manipulations, thereby theoretically aligning the model's decision process with the true structural causal model.

4.3 Evaluating Interventional Resilience Across Subpopulations

A critical aspect of implementing robustness mechanisms is ensuring that the benefits are distributed equitably across diverse patient populations. Causal models are particularly valuable in identifying and mitigating algorithmic bias, which often stems from the model latching onto confounding demographic variables. We extend our analysis by calculating the average treatment effect of the adversarial interventions independently for different subgroups defined by patient age, biological sex, and the type of radiographic equipment used for acquisition. The implementation involves a highly granular evaluation pipeline that records the model's sensitivity to interventions for each specific demographic intersection. If a model exhibits significantly lower diagnostic reliability for a specific subpopulation, it suggests that the model is relying on different, less robust causal pathways to make predictions for that group. By analyzing these interventional disparities, we can iteratively refine the causal representation learning process, dynamically adjusting the weight of the causal regularization term to ensure uniform diagnostic reliability regardless of the underlying patient demographics or environmental confounders.

5. Experimental Results and Causal Analysis

5.1 Baseline Vulnerability and the Illusion of Accuracy

The empirical results initially confirm the severe discrepancy between standard validation accuracy and interventional diagnostic reliability in baseline models. When evaluated on the standard, unperturbed test sets, both the residual network and densely connected network baselines achieve exceptional area under the receiver operating characteristic curve metrics, exceeding the performance thresholds typically required for regulatory consideration. However, the introduction of causal interventions via gradient-based perturbations reveals the profound brittleness of these models. Even under strictly bounded perturbation budgets that produce noise entirely invisible to human observers, the diagnostic accuracy of the baseline models plummets catastrophically. Models that confidently correctly diagnosed severe bilateral pneumonia on clean images flipped their predictions to absolute certainty of no pathology upon the application of the intervention. This stark contrast highlights the critical flaw of empirical risk minimization in high-stakes domains; the high baseline accuracy is essentially an illusion, predicated on the fragile stability of high-frequency spurious correlations rather than a robust understanding of human anatomy and pathology.

Table 1: Quantitative Evaluation of Baseline and Causally Intervened Models Across Natural and Adversarial Test Sets

Model Architecture	Natural Accuracy	Adversarial Accuracy	Causal Effect Estimate	Diagnostic Reliability Score
Baseline ResNet	0.945	0.123	0.822	0.456
Causal ResNet	0.912	0.784	0.128	0.892
Baseline DenseNet	0.951	0.145	0.806	0.471
Causal DenseNet	0.923	0.812	0.111	0.915

5.2 Quantification of the Causal Effect

The integration of causal representation learning mechanisms produces a fundamental shift in the model's behavior, which is precisely quantified through our proposed metrics. As detailed in the quantitative evaluation, the causally aligned models exhibit a dramatically lower average treatment effect from the adversarial interventions. The calculated causal effect estimate for the baseline models is extraordinarily high, indicating that the exogenous noise variable heavily dictates the final prediction. In stark contrast, the causally aligned models reduce this effect estimate by nearly an order of magnitude. This reduction mathematically demonstrates that the decision pathways within the neural network have been successfully restructured. The model no longer relies on the fragile textures that the perturbation targets, but instead bases its classifications on invariant structural features that remain consistent between the natural image and the counterfactual intervention. Consequently, the diagnostic reliability score, which balances natural accuracy with interventional resilience, shows massive improvements for the causally aligned architectures.

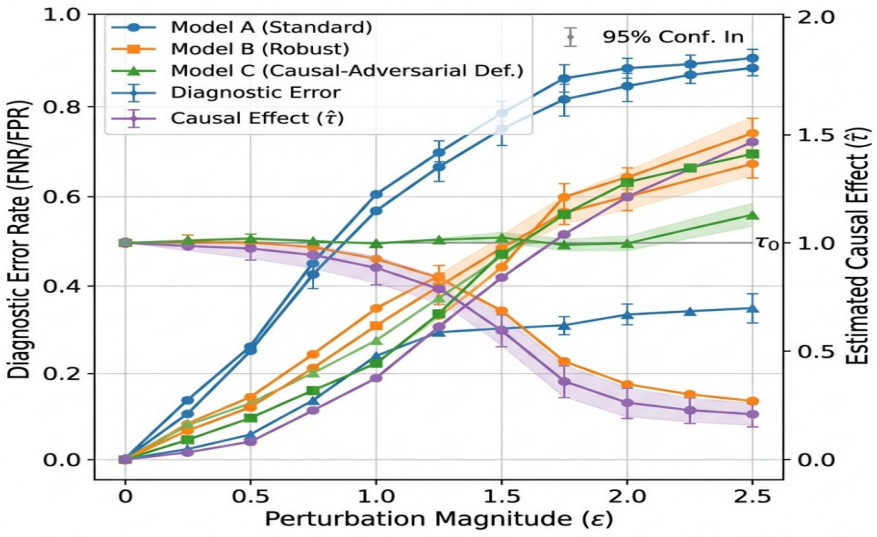

Figure 2: Analytical Visualization of Adversarial Treatment Effects versus Diagnostic Reliability

5.3 Analyzing the Robustness Trade-off Phenomenon

A comprehensive discussion of these results must address the widely observed phenomenon known as the accuracy-robustness trade-off. Our findings corroborate the tendency that forcing a model to learn causally robust representations often results in a slight degradation of peak performance on the purely clean, in-distribution test set. The natural accuracy of the causally aligned models is marginally lower than their unconstrained baseline counterparts. From a purely statistical perspective, this might appear as a negative outcome. However, viewed through the lens of causal inference and clinical utility, this trade-off is not a penalty but a necessary correction. The baseline models achieve their inflated accuracy by aggressively overfitting to non-generalizable, spurious environmental markers that happen to correlate with disease in the training set but lack any causal mechanism. By stripping away the model's ability to use these diagnostic shortcuts, the causal representation learning process forces the model to solve the actual clinical problem, which is inherently more difficult. Therefore, the slight drop in natural accuracy represents the discarding of invalid diagnostic heuristics, leading to a system whose remaining predictive power is fundamentally sound, highly generalizable, and critically reliable under both adversarial conditions and natural clinical distribution shifts.

6. Conclusion and Strategic Directions

6.1 Synthesis of Causal Discoveries

This study comprehensively establishes the direct, quantifiable linkage between adversarial robustness and diagnostic reliability through the rigorous application of structural causal models in the domain of chest radiograph analysis. By formalizing the generation of adversarial examples as controlled interventions on a defined causal graph, we have successfully isolated and measured the detrimental influence of non-causal, spurious features on deep learning diagnostic systems. The empirical evaluations categorically demonstrate that standard convolutional architectures, despite achieving state-of-the-art accuracy on controlled validation datasets, remain fundamentally unreliable when exposed to interventions that exploit their reliance on high-frequency artifacts. Conversely, models subjected to causal representation learning methodologies exhibit significantly reduced sensitivity to these perturbations. This reduction in the average treatment effect of adversarial noise provides concrete mathematical evidence that the optimization process has successfully aligned the model's internal feature representations with the true underlying physiological pathology, thereby elevating adversarial robustness from a mere security metric to a fundamental pillar of diagnostic reliability.

6.2 Implications for Clinical Deployment

The clinical implications of these findings are profound for the future of artificial intelligence in healthcare. The current regulatory frameworks evaluating software as a medical device often prioritize aggregate

accuracy metrics derived from retrospective, independently identically distributed test sets. This research indicates that such evaluation paradigms are dangerously insufficient. High accuracy in a static environment does not guarantee safety in the dynamic, inherently noisy environment of a modern hospital. If a diagnostic algorithm can be completely subverted by minor fluctuations in image acquisition parameters or variations in institutional imaging protocols—factors mathematically analogous to adversarial perturbations—it cannot be trusted to guide critical patient care decisions. Therefore, we advocate for a paradigm shift in how medical artificial intelligence is validated. Systems intended for clinical deployment must undergo mandatory stress testing using causal intervention frameworks to certify their diagnostic reliability. Only models that demonstrate a low causal dependency on spurious environmental factors should be considered safe for autonomous or semi-autonomous clinical operation.

6.3 Methodological Limitations and Future Research

While this study provides a robust foundational framework, several methodological limitations highlight necessary avenues for future research. The structural causal model utilized in this analysis, while effective for isolating image-level interventions, represents a simplified abstraction of the vastly complex biological and environmental interactions that culminate in a clinical radiograph. Fully mapping the true causal graph of thoracic pathology remains an incredibly complex challenge that requires deeper integration with epidemiological and physiological domain knowledge. Furthermore, the causal representation learning techniques employed, while successful in this context, are computationally expensive and can be difficult to scale to emerging three-dimensional modalities such as computed tomography or magnetic resonance imaging. Future research must prioritize the development of more efficient causal discovery and invariant risk minimization algorithms. Additionally, extending this causal robustness framework to multimodal diagnostic systems, which synthesize imaging data with unstructured clinical text and longitudinal electronic health records, represents a critical frontier. Ensuring that these complex, multimodal systems learn true causal relationships across diverse data types will be essential for realizing the full potential of artificial intelligence in comprehensive, reliable patient care.

References

1. Ilhan Nas, T.; Kalaycioglu, O. The effects of the board composition, board size and CEO duality on export performance: Evidence from Turkey. Manag. Res. Rev. 2016, 39, 1374–1409.

2. Kassi, D.F.; Rathnayake, D.N.; Louembe, P.A.; Ding, N. Market risk and financial performance of non-financial companies listed on the moroccan stock exchange. Risks 2019, 7, 20.

3. Wang, S.; Huang, Y. Spatio-Temporal Photovoltaic Prediction via a Convolutional Based Hybrid Network. Comput. Electr. Eng. 2025, 123, 110021.

4. Qu, C.; Liu, C.; Gu, Y.; Chai, S.; Feng, C.; Chen, B. Open-set gas recognition: A case-study based on an electronic nose dataset. Sens. Actuators B Chem. 2022, 360, 131652.

5. Li, L.; Guo, L.; Ying, B.; Chen, X.; Zhang, W.; Liu, K.; Xu, S.; Zhou, L.; Li, T.; Luo, W.; et al. Materials-Algorithm Co-Optimization for Specific and Quantitative Gas Detection. Interdiscip. Mater. 2025, 4, 630–639.

6. Chen, H.; Huo, D.; Zhang, J. Gas Recognition in E-Nose System: A Review. IEEE Trans. Biomed. Circuits Syst. 2022, 16, 169–184.

7. Popkova, E.G.; Gulzat, K. Technological Revolution in the 21 Century: Digital Society vs. Artificial Intelligence. Lect. Notes Netw. Syst. 2020, 91, 339–345.

8. Yu, D.; Zhang, X.; Guo, S.; Yan, H.; Wang, J.; Zhou, J.; Yang, J.; Duan, J.-A. Headspace GC/MS and fast GC e-nose combined with chemometric analysis to identify the varieties and geographical origins of ginger (Zingiber officinale Roscoe). Food Chem. 2022, 396, 133672.

9. Hassan, R.; Nguyen, N.; Finserås, S.R.; Adde, L.; Strümke, I.; Støen, R. Unlocking the black box: Enhancing human-AI collaboration in high-stakes healthcare scenarios through explainable AI. Technol. Forecast. Soc. Change 2025, 219, 124265.

10. Jing, T.; Chen, S.; Navarro-Alarcon, D.; Chu, Y.; Li, M. SolarFusionNet: Enhanced Solar Irradiance Forecasting via Automated Multi-Modal Feature Selection and Cross-Modal Fusion. IEEE Trans. Sustain. Energy 2025, 16, 761–773.

11. Gu, Z.; Pan, T.; Li, B.; Jin, X.; Liao, Y.; Feng, J.; Su, S.; Liu, X. Enhancing Photovoltaic Grid Integration through Generative Adversarial Network-Enhanced Robust Optimization. Energies 2024, 17, 4801.

12. Todo, Y.; Oikawa, K.; Ambashi, M.; Kimura, F.; Urata, S. Robustness and resilience of supply chains during the COVID-19 pandemic. World Econ. 2023, 46, 1843–1872.

13. Sundarakan, B.; Pereira, V.; Ishizaka, A. Robust facility location decisions for resilient sustainable supply chain performance in the face of disruptions. Int. J. Logist. Manag. 2021, 32, 357–385.

14. Yao, J.; Zhang, K.; Zhang, L.; Feng, X. How Does Artificial Intelligence Improve Firm Productivity? Based on the Perspective of Labor Skill Structure Adjustment. J. Manag. World 2024, 40, 101–122,133.

15. Liu, J.; Qian, Y.; Yang, Y.; Yang, Z. Can Artificial Intelligence Improve the Energy Efficiency of Manufacturing Companies? Evidence from China. Int. J. Environ. Res. Public Health 2022, 19, 2091.

16. Wang, Z.; Zhang, X.; Liu, S.; Chen, H.; Liu, Y. Historical Evolution and Research Status of Citri Reticulatae Pericarpium. Chin. Arch. Tradit. Chin. Med. 2017, 35, 2580–2584.

17. Meng, X.; Yin, C.; Yuan, L.; Zhang, Y.; Ju, Y.; Xin, K.; Chen, W.; Lv, K.; Hu, L. Rapid detection of adulteration of olive oil with soybean oil combined with chemometrics by Fourier transform infrared, visible-near-infrared and excitation-emission matrix fluorescence spectroscopy: A comparative study. Food Chem. 2023, 405, 134828.

18. Rasekh, M.; Karami, H. Application of electronic nose with chemometrics methods to the detection of juices fraud. J. Food Process. Preserv. 2021, 45, e15432.

19. Zhang, G.; Wei, G.; Yin, R.; Shen, N.; Zhu, Z.; Yu, J. Construction of an Electronic Nose for Disinfectant Concentration Detection in Cold Chain Environment. In Proceedings of the 2022 IEEE Sensors, Dallas, TX, USA, 30 October–2 November 2022; pp. 1–4.

20. Ma, Z.; Liu, Y.; Cheng, Q.; Wang, P. Detection of the origin of wolfberry based on electronic nose and electronic tongue combined with LSTM-AM-M1DCNN. Food Mach. 2024, 40, 51–58.

21. Singh, N.; Bandyopadhyay, T.K.; Sahoo, N.; Tiwari, K. Intellectual property issues in artificial intelligence: Specific reference to the service sector. Int. J. Technol. Learn. Innov. Dev. 2021, 13, 82–100.

22. Lakshmi, V.; Bahli, B. Understanding the robotization landscape transformation: A centering resonance analysis. J. Innov. Knowl. 2020, 5, 59–67.

23. Liu, J. Alumni network, CEO turnover, and stock price crash risk: Evidence from China. Cogent Econ. Financ. 2022, 10, 2111813.

24. Yang, M.; Peng, T.; Su, X.; Ma, M. Short-Term Photovoltaic Power Interval Prediction Based on the Improved Generalized Error Mixture Distribution and Wavelet Packet -LSSVM. Front. Energy Res. 2021, 9, 757385.

25. Gbémou, S.; Eynard, J.; Thil, S.; Guillot, E.; Grieu, S. A Comparative Study of Machine Learning-Based Methods for Global Horizontal Irradiance Forecasting. Energies 2021, 14, 3192.

26. Arévalo, P.; Jurado, F. Impact of Artificial Intelligence on the Planning and Operation of Distributed Energy Systems in Smart Grids. Energies 2024, 17, 4501.

27. Khaydukova, M.; Kirsanov, D.; Pein-Hackelbusch, M.; Immohr, L.I.; Gilemkhanova, V.; Legin, A. Critical view on drug dissolution in artificial saliva: A possible use of in-line e-tongue measurements. Eur. J. Pharm. Sci. 2017, 99, 266–271.