BIOPRESS JOURNAL OF COMPUTATIONAL LIFE SCIENCES, VOL. 2, ISSUE. 2 (2026)

Artificial Intelligence in Modern Diagnostic Systems: From Explainable Models to Clinical and Industrial Applications

Sandipan Chatterjee, Rupsa Chakraborty, Soumyadeep Mukherjee, Prasun Chowdhury
Department of ECE, St. Thomas' College of Engineering & Technology
Kolkata, India

Corresponding author Email: ec23.sandipan.chatterjee@stcet.ac.in

Doi: 10.5281/zenodo.18834944
Received: 10 February 2026
Accepted: 18 February 2026

Abstract

Artificial Intelligence (AI) has emerged as the central engine of next-generation diagnostic systems, reshaping medical imaging, precision agriculture, and intelligent industrial maintenance. This review presents a unified technical perspective on computer-aided diagnosis (CAD) empowered by deep learning and Explainable AI (XAI). We synthesize methodologies ranging from convolutional neural networks (CNN), transfer learning, hybrid multimodal fusion, to generative language assistants. Mathematical foundations of learning, optimization, and attribution are detailed alongside pseudocode for reproducible pipelines. Comparative tables analyze algorithms across accuracy, interpretability, computational cost, and regulatory readiness.

Keywords: CAD, XAI, CNN, Grad-CAM, SHAP, Medical Imaging, Precision Agriculture, Generative AI, Edge Intelligence

1. Introduction

AI-driven diagnostic systems have shifted from rule-based decision trees toward data-driven representation learning, enabling robust pattern recognition in high-dimensional signals such as MRI, CT, X-ray, spectroscopy, and sensor logs [6–9]. In radiology, deep learning has shown strong performance across detection, classification, and segmentation tasks, but clinical adoption has been slowed by model opacity, dataset shift, and governance constraints [7,10,11]. These barriers are particularly salient for high-stakes decisions where clinicians must justify actions, trace evidence, and communicate uncertainty [1,2,10,12]. As a result, Explainable AI (XAI) has become a central requirement for diagnostic AI—both to support human interpretability and to meet emerging expectations for transparency and auditability [3,4,7,12].

Recent work demonstrates that explainability can meaningfully affect real-world acceptance: transparent decision-support systems combining performance with human-aligned explanations can increase clinician confidence and reduce workflow friction [2,10,13]. Meanwhile, medical imaging AI has exposed fairness risks: deep models can infer sensitive attributes such as self-reported race from medical images at high accuracy, creating pathways for unintended bias if not carefully audited [14–16]. These findings further motivate systematic evaluation, reporting, and monitoring guidance such as CONSORT-AI, SPIRIT-AI, and TRIPOD+AI for AI-based clinical studies [17–20].

Beyond healthcare, similar diagnostic paradigms have expanded into agriculture and industrial maintenance. Vision models have been widely adopted for plant disease detection, where data scarcity and environmental variation demand transfer learning and robust augmentation [21,22]. In coconut disease diagnostics, public datasets and CNN pipelines have accelerated research and tool development for farmers [23–25]. In industrial contexts, vehicle diagnostics is increasingly augmented by AI assistants that interpret trouble codes and support troubleshooting in natural language, with embedded/edge deployments emerging as practical architectures [5].

AI-based diagnostics can be formulated as a supervised mapping

$$f_{\theta}: \mathcal{X} \rightarrow \mathcal{Y}$$

where $\mathcal{X}$ is the space of medical images, sensor signals, or clinical records and $\mathcal{Y}$ is the diagnostic label set [6–9]. The learning objective minimizes empirical risk:

$$\theta^{\backslash *} = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(f_{\theta}(x_i), y_i) + \lambda \parallel \theta \parallel_2^2$$

where $\mathcal{L}$ is cross-entropy or focal loss for imbalanced disease classes [7,10].

Pseudocode – Generic Diagnostic Learning

```
Input: Dataset D={(x_i,y_i)}
Initialize θ
for epoch =1..E:
    for minibatch B:
        ŷ = fθ(x)
        loss = CrossEntropy(ŷ,y) + $\lambda\|\theta\|^2$
        θ ← θ - η∇θ loss
return model fθ
```

Such formulation underlies brain tumor CAD [1], DSS systems [2], and plant disease tools [24]. However, model opacity violates clinical requirements for justification [3,4,12]. Therefore, XAI generates an explanation operator

$$g(f_{\theta}, x) \rightarrow e$$

producing saliency map e aligned with human reasoning [5].

2. Core AI Models for Diagnostics

2.1 CNNs, Transfer Learning, and Self-Configuring Pipelines

CNNs remain foundational for image diagnostics because convolutional feature hierarchies capture edges, textures, and semantic structures in a stable, translation-aware manner [6,7]. Transfer learning (e.g., VGG/ResNet/Inception family models) reduces annotation requirements by reusing general visual representations [6,21]. For segmentation, self-configuring frameworks like nnU-Net automatically adapt preprocessing, architecture, and training recipes to new medical tasks, setting strong baselines in biomedical image segmentation [26]. This has influenced "AutoML-like" best practices in clinical AI where reproducibility and standardized baselines are valued [10,26].

A convolution layer computes

$$h_{i,j,k} = \sigma\left(\sum\nolimits_{m,n,c} W_{m,n,c,k}\ x_{i+m,j+n,c} + b_k\right)$$

where σ is ReLU or Swish [1]. Feature maps cascade through pooling:

$$p_{i,j} = \max_{m,n \in \Omega} h_{i+m,j+n}$$

Transfer learning reuses weights W^{pre} from ImageNet:

$$W = W^{pre} + \Delta W$$

minimizing fine-tuning cost for limited medical data [6,21].

Pseudocode – Transfer Training

```
Model ← Pretrained(VGG16)
Freeze early layers
Replace FC → Softmax(K)
Train last layers on medical set
Unfreeze + low LR fine-tune
```

2.2 Transformers in Medical Image Analysis

Transformers (and hybrid CNN-Transformer models) have become increasingly influential in medical imaging pipelines, offering improved global-context modeling for segmentation and classification [27,28]. Surveys covering 100+ transformer papers in medical imaging outline applications in segmentation, reconstruction, report generation, and multimodal learning [27]. Swin-based and 3D transformer variants have shown strong results in volumetric segmentation and whole-organ analysis [28,29].

3. Explainable AI Techniques in Diagnostic Workflows

3.1 Attribution and Post-hoc Explanations

Post-hoc explainers such as LIME, SHAP, and Grad-CAM are widely used because they can be applied to complex models without requiring a fully interpretable architecture [3–5]. Grad-CAM produces class-discriminative saliency maps by weighting final-layer activations with gradients, enabling clinicians to verify whether a model focuses on plausible anatomical regions in imaging tasks [5]. SHAP uses Shapley-value principles to assign feature contributions to predictions, supporting tabular clinical risk models and multimodal DSS [4,10]. LIME fits local surrogate models around individual predictions, offering instance-level feature explanations to clinicians and auditors [3].

Large-scale surveys of XAI in medical imaging describe common pitfalls: explanations may be unstable under perturbations, may reflect spurious correlations, and can create “illusion of transparency” if not validated against clinical reasoning [7,11,30]. Recent work also emphasizes the need for explanation evaluation beyond visual appeal, using human-grounded metrics, faithfulness checks, and robustness tests [10,11,30].

3.1.1 Grad-CAM

For class c importance weights:

$$\alpha_k^c = \frac{1}{Z} \sum_{i,j} \frac{\partial y^c}{\partial A_{ij}^k}$$

Heatmap:

$$L_{GradCAM}^c = ReLU\left(\sum_k \alpha_k^c A^k \right) \text{ [5]}$$

Pseudocode – GradCAM

```
Forward x → activations A
Backward class c → gradients
α ← GlobalAvgPool(gradients)
Heat ← ReLU(Σ α_k A_k)
Overlay Heat on image
```

3.1.2 SHAP

Feature contribution:

$$\phi_i = \sum_{S \subseteq F \setminus i} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup i) - f(S)] \quad [4]$$

3.1.3 LIME

Local surrogate:

$$\min_{g} \ \mathcal{L}(f, g, \pi_x) + \Omega(g) \qquad [3]$$

3.2 Why "Trust" Requires More Than Heatmaps

Trust in medical AI depends on (i) performance under domain shift, (ii) explanation reliability, and (iii) institutional governance [10,12,18]. Fairness concerns are nontrivial: evidence shows AI can detect demographic attributes from images even when radiologists cannot, complicating efforts to ensure equity across populations [14–16]. Furthermore, fairness can have "limits" in real-world deployment because demographic proxies and hidden signals may persist across data sources [15,16]. These results motivate rigorous bias audits, cross-site validation, and transparent documentation [10,12,19,20].

4. Medical Diagnostics: Brain Tumors and Radiology

4.1 Brain Tumor CAD with XAI

Brain tumor diagnosis benefits from CAD systems that improve sensitivity and speed while reducing inter-observer variability [1,31]. XAI-enhanced workflows overlay Grad-CAM saliency maps on MRI to highlight regions driving classification decisions, improving interpretability and clinician confidence [1,5]. In practice, brain tumor pipelines commonly include: preprocessing (skull stripping/normalization), CNN feature extraction, classification or segmentation, and explanation generation [1,26,31]. Modern baselines often compare VGG/ResNet backbones and can incorporate volumetric segmentation via nnU-Net or transformer models for higher-fidelity localization [26–29,31].

4.2 Radiology Workflow Automation and Validation

Radiology has seen growing integration of AI in triage, detection, and reporting augmentation [6,7]. Clinical reviews emphasize that while AI can detect subtle patterns, operational safety requires continuous monitoring, calibration, and reporting discipline to avoid silent failure modes under shift [6,10]. Model governance practices align with GMLP guiding principles and international medical-device recommendations that stress lifecycle management, data quality, and post-deployment monitoring [32,33]. Reporting standards such as CONSORT-AI, SPIRIT-AI, and TRIPOD+AI further support transparent study communication and enable reproducible evidence synthesis [17–20].

4.3 Preprocessing

Intensity normalization:

$$x' = \frac{x - \mu}{\sigma}$$

Augmentation transform T:

$$\tilde{x} = T(x) = R_\theta S_\gamma F$$

4.4 Classification Model

Softmax output:

$$p_k = \frac{e^{z_k}}{\sum_j e^{z_j}}$$

Loss:

$$\mathcal{L} = -\sum_k y_k \log p_k$$

4.5 Full Pseudocode

```
for MRI in dataset:
    x ← skull_strip(MRI)
    x ← normalize(x)
    x ← augment(x)

Model ← VGG16+Swish
Train with Adam
for test sample:
    ŷ ← Model(x)
    Heat ← GradCAM(Model,x)
```

5. Decision Support Systems and Multimodal Diagnostics

Transparent AI-driven DSS can combine tabular EHR, lab values, and imaging features in hybrid architectures (e.g., CNN encoders + RNN/transformer clinical encoders) [2,27]. Such systems often use SHAP or related attributions to show the relative influence of clinical variables, and LIME for localized explanations of individual risk scores [2–4]. However, multimodal fusion introduces new failure modes: missingness patterns, modality imbalance, and confounding can distort both predictions and explanations [10,12]. Therefore, best practices increasingly emphasize external validation, subgroup reporting, and careful documentation of cohort selection [10,18–20].

Fusion of image features $v$and clinical vector c:

$$h = \tanh(W_v v + W_c c + b)$$
$$p = \sigma(Wh)$$

SHAP explanation:

$$Explain = \{\phi_{age}, \phi_{BP}, \phi_{lesion}\}$$

Pseudocode

```
v ← CNN(image)
c ← EHR_vector
h ← concat(v,c)
p ← classifier(h)
shap ← SHAP(p)
```

6. Agricultural Diagnostics: Plant and Coconut Disease Detection

Deep learning for plant disease detection has expanded rapidly because image-based symptom classification maps well to CNN feature learning [21,22]. Surveys highlight key practical constraints: field conditions (lighting, occlusion), class imbalance, and domain shift between lab datasets and real farms [21,22]. Explainable AI is increasingly used to highlight symptomatic regions and support farmer trust, especially in decision tools where treatment guidance is provided [24,34,35]. Public datasets have accelerated coconut disease research, including multi-class disease corpora for bud rot, leaf rot, stem bleeding and other conditions [23,25]. Recent coconut-specific modeling studies report high accuracy with CNN families and improved robustness via multi-scale feature extraction and optimization [23–25].

Triage score:

$$s = \beta_1 p_{nodule} + \beta_2 p_{bleed}$$

Queue ordering:

$$order = argsort(-s)$$

Pseudocode

```
for study in PACS:
    p ← AI(study)
    priority ← β·p
sort by priority
auto-report via NLP
```

7. Industrial Diagnostics: Vehicles and Embedded AI Assistants

Vehicle diagnostics traditionally relies on trouble code readouts and technician expertise; AI assistants can translate codes into likely causes and recommended actions using structured prompts and retrieval strategies [5,36]. Embedded diagnostic architectures using Wi-Fi enabled OBD/CAN modules combined with cloud-based generative AI services can enable autonomous or semi-autonomous troubleshooting while maintaining a conversational interface [5]. However, safety requires careful constraints: hallucination risk, grounding in verified manuals, and secure device communication are necessary for responsible deployment [36,37]. These concerns mirror those in healthcare, where LLMs show promise in decision support but require strict evaluation, guardrails, and workflow integration to prevent errors [38–40].

7.1 Inception Feature Fusion

$$F = [f_{1\times1}, f_{3\times3}, f_{5\times5}]$$

Prediction:

$$p = \mathrm{softmax}(WF)$$

Pseudocode

```
img ← upload
x ← preprocess(299)
p ← Inception(x)
Heat ← GradCAM(x)
```

8. Privacy-Preserving and Federated Diagnostics

Privacy constraints limit centralized data pooling in both hospitals and industrial fleets. Federated learning (FL) is increasingly explored for medical imaging because it supports collaborative training without direct data sharing, while still facing challenges such as heterogeneity, communication cost, and privacy leakage through gradients [41–44]. Recent surveys summarize FL methods for medical imaging and discuss practical deployment patterns (cross-silo FL, secure aggregation, personalization layers) [41–44]. FL aligns with governance demands by reducing data movement, but it does not eliminate the need for bias audits, explanation validation, and post-market monitoring [10,32,33,41].

Bayesian reasoning over DTCs:

$$P(cause \mid dtc) \propto P(dtc \mid cause)P(cause)$$

LLM prompt:

```
context = {VIN, ECU, DTC}
response ← LLM(context)
```

9. Reporting, Regulation, and Deployment Readiness

Clinical AI credibility depends on transparent reporting and regulatory-aligned evidence generation. CONSORT-AI and SPIRIT-AI extend standard trial reporting and protocol guidance to AI interventions, emphasizing reproducible description of data, algorithm behavior, human-AI interaction, and failure modes [17,18]. TRIPOD+AI provides updated reporting guidance for prediction models built with ML, enabling better comparisons and meta-analyses [19,20]. Device lifecycle guidance such as GMLP and international medical-device principles stress data quality, continuous monitoring, and change management for ML-enabled devices [32,33]. Together, these frameworks help bridge the gap between “high performance in retrospective datasets” and “safe performance in real-world workflows” [10,12,32,33].

Global update:

$$w_{t+1} = \sum_k \frac{n_k}{N} w_k$$

Pseudocode

```
for round t:
  each hospital k:
      wk ← train_local(w)
```

aggregate wk

10. Evaluation Metrics

Accuracy

$$Acc = \frac{TP + TN}{N}$$

F1

$$F1 = \frac{2PR}{P + R}$$

Faithfulness of XAI:

$$Faith = \operatorname{corr}(e, \Delta f)$$

11. End-to-End Unified Algorithm

Input: multi-domain data

1. Preprocess

2. Train CNN/Transformer

3. Validate external

4. Generate XAI

5. Human review loop

6. Deploy with monitoring

12. Comparative Analysis

Table 1 – Algorithm Comparison

Domain	Model	Accuracy		XAI	Strengths	Limitations
Brain MRI	VGG16+GradCAM	~95%		High	Visual rationale	Data bias
DSS	Hybrid CNN-RNN	94.8%		SHAP/LIME	Multimodal	Integration
Coconut	InceptionV3	92%		Heatmap	Field use	Lighting
Vehicle	LLM	Qualitative		Text	Reasoning	Hallucination
Radiology	CNN/Triage	Workflow		Visual	Prioritization	Privacy

Table 2 – Explainability Techniques

Method	Type	Output	Use Case
Grad-CAM	Visual	Heatmap	Imaging
SHAP	Feature	Scores	EHR
LIME	Local	Rules	DSS
Attention	Intrinsic	Weights	Transformers

13. Conclusion with Formal View

Diagnostic AI is an operator chain:

$$x \xrightarrow{f_\theta} y \xrightarrow{g} e \xrightarrow{H} decision$$

where H is human cognition. This synergy ensures safe adoption [10,12,32].

AI diagnostics have matured into practical systems spanning clinics, farms, and vehicles. Evidence from multiple studies confirms that explainability is the key enabler of trust and adoption . The next decade will witness deeper human-AI symbiosis guided by transparency and ethics.

14. Challenges

- Dataset shift
- Bias & fairness
- Privacy
- LLM hallucination
- Clinical validation

15. Future Directions

- Federated learning across hospitals
- Causal XAI
- Edge TinyML for farms
- Guard-railed generative assistants

References

[1] S. T. Padmapriya and M. S. Gayathri Devi, “Computer-Aided Diagnostic System for Brain Tumor Classification using Explainable AI,” IEEE, 2024.
[2] A. M. Annamalai et al., “Revolutionizing Medical Diagnostics with Transparent AI-Driven Decision Support Systems,” IEEE, 2024.
[3] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why Should I Trust You? Explaining the Predictions of Any Classifier,” KDD, 2016.
[4] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions (SHAP),” NeurIPS, 2017.
[5] R. R. Selvaraju et al., “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” ICCV, 2017.

[6] P. M. Cheng et al., “Deep Learning: An Update for Radiologists,” Radiographics, 2021.
[7] B. H. M. van der Velden et al., “Explainable AI in Deep Learning-based Medical Image Analysis: A Survey,” Med Image Anal, 2022.
[8] C. L. Kok et al., “Enhancing Diagnostic Accuracy: The Role of AI in Advanced Radiological Imaging,” IEEE, 2024.
[9] D. Bhati et al., “A Survey on Explainable AI Techniques for Visualizing Deep Learning Models in Medical Imaging,” J. Imaging, 2024.
[10] S. H. Park, “Reporting Guidelines for Artificial Intelligence Studies in Radiology (incl. TRIPOD+AI, CLAIM),” Korean J Radiol, 2024.
[11] M. T. Zamir et al., “Explainable AI-Driven Analysis of Radiology Reports…” 2025.
[12] F. Ahmed et al., “Explainable AI in Medical Imaging,” BMC Med Imaging, 2026 (ahead listing includes 2025 items).
[13] A. P. L. Martindale et al., “Concordance of RCTs for AI Interventions with CONSORT-AI,” Nat Commun, 2024.
[14] J. W. Gichoya et al., “AI recognition of patient race in medical imaging,” Lancet Digit Health, 2022.
[15] Y. Yang et al., “The limits of fair medical imaging AI in real-world settings,” Nat Med, 2024.
[16] J. L. Burns et al., “Ability of AI to identify race from medical images…” 2023.
[17] X. Liu et al., “CONSORT-AI extension,” Nat Med, 2020.
[18] S. Cruz Rivera et al., “SPIRIT-AI extension,” Nat Med, 2020.
[19] G. S. Collins et al., “TRIPOD+AI statement,” BMJ, 2024.
[20] TRIPOD+AI (EQUATOR summary page), 2024.
[21] A. Ahmad et al., “Survey on deep learning for plant disease diagnosis and management,” Smart Agric Tech, 2023.
[22] Plant disease detection survey PDF, 2023.
[23] S. Thite et al., “Coconut (Cocos nucifera) Tree Disease Dataset,” Data in Brief, 2023.
[24] S. Santhi et al., “Cocoscan: AI-Powered Precision Diagnostics for Coconut Leaf Disease,” IEEE, 2025.
[25] X. Huang et al., “Disease Infection Classification in Coconut Tree Based on Deep Learning,” Processes, 2025.
[26] F. Isensee et al., “nnU-Net: a self-configuring method for biomedical image segmentation,” Nat Methods, 2021.
[27] F. Shamshad et al., “Transformers in medical imaging: A survey,” Med Image Anal, 2023.
[28] Y. Cai et al., “Swin Unet3D…” 2023.
[29] C. Wei et al., “High-Resolution Swin Transformer for Medical Image Segmentation,” Sensors, 2023.
[30] R. Santos et al., “Grad-CAM analysis and receptive field impact,” 2025.
[31] BraTS/brain tumor segmentation benchmark background (via nnU-Net and related literature), 2021–2024.
[32] FDA/Regulators, “Good Machine Learning Practice (GMLP) Guiding Principles,” updated FDA page (and linked docs), 2025.
[33] IMDRF, “GMLP for medical device development: Guiding principles,” 2025.
[34] M. T. Ahmed et al., “Systematic review of XAI for agriculture/food analytics,” 2025.
[35] Agriculture 5.0 and Explainable AI scoping review, 2023.
[36] V. Ranchev et al., “Integration of Generative AI for Intelligent Diagnostic of Vehicles,” IEEE, 2024.
[37] Security considerations for embedded Wi-Fi/OBD diagnostics (overview from vehicle diagnostic architecture paper), 2024.
[38] Y. Artsi et al., “Large language models in real-world clinical workflows,” Frontiers Digital Health, 2025.
[39] J. Vrdoljak et al., “Review of large language models in medical education and clinical decision support,” 2025.
[40] J. Li, “Large language models-powered clinical decision support,” 2025.
[41] H. Guan et al., “Federated learning for medical image analysis: A survey,” 2024.
[42] F. R. da Silva et al., “Federated Learning in Medical Image Analysis,” Electronics, 2023.
[43] G. Choi et al., “Survey of medical applications of federated learning,” 2024.
[44] “Federated learning meets medical image analysis,” 2024 review/monograph.
[45] M. S. Krishna et al., “Plant Leaf Disease Detection Using Deep Learning…” MDPI, 2025.