International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(9s) 2026 pp. 841-848
https://doi.org/10.70917/ijcisim-2026-3506
Article

Smart Healthcare Systems: AI-Driven Innovations for Disease Prediction, Diagnosis, and Public Health Management

Ruchi Gupta[1], Sandeep Gupta[2], S. Malathi[4] , Lonavath Srinivas Naik[5] , D. Swapna[6]

[1]Department Of Information Technology, Kuala Lumpur University of Science & Technology, Malaysia & Ajay Kumar Garg Engineering College, Ghaziabad , Ghaziabad, Uttar Pradesh, guptaruchi@akgec.ac.in
[2]Department of Computer Science & Engineering, Sharda University, Greater Noida & Faculty of Engineering & Technology KLUST, Malaysia,Gautam Budh Nagar, Greater Noida, Uttar Pradesh, guptasandeep1093@gmail.com
[4]Mathematics, K. Ramakrishnan College of Technology, Tiruchirappalli Samayapuram, Tamil Nadu, malamathematics@gmail.com
[5]Department of Mechanical Engineering. School of Engineering, Anurag University, Medchal Malkajgiri, Hyderabad, Telangana, srinivasnaikmech@anurag.edu.in
[6]Mechanical Engineering, R.V.R & J.C. College of Engineering, Guntur, Andhra Pradesh, dhulipallaswapna@gmail.com

Abstract: The convergence of artificial intelligence (AI), the Internet of Medical Things (IoMT), and big data analytics has transformed healthcare delivery from a reactive to a predictive and preventive paradigm. This article examines the role of AI-driven smart healthcare systems in disease prediction, clinical diagnosis, and public health surveillance. Machine learning (ML) and deep learning (DL) architectures—including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, Random Forest, and XGBoost—are analyzed for their diagnostic performance across cardiovascular disease, diabetes, cancer, and infectious disease outbreaks such as COVID-19. Comparative accuracy benchmarks, sensitivity/specificity metrics, and real-world deployment case studies are presented in tabular form. The article further discusses federated learning for privacy-preserving diagnostics, wearable-based continuous monitoring, and AI-enabled epidemic forecasting models used by public health agencies. Challenges related to data heterogeneity, algorithmic bias, interpretability, and regulatory compliance are critically discussed, followed by future research directions toward explainable AI (XAI) and edge-computing-based healthcare infrastructure.

Keywords: —Artificial Intelligence, Smart Healthcare, Disease Prediction, Deep Learning, Public Health Surveillance, IoMT, Federated Learning

1. INTRODUCTION

Healthcare systems worldwide are under increasing strain due to aging populations, rising chronic disease burden, and resource-constrained public health infrastructure. The World Health Organization estimates that noncommunicable diseases (NCDs) such as cardiovascular disease, cancer, diabetes, and chronic respiratory disease account for approximately 74% of global deaths annually [1]. Traditional diagnostic workflows, which rely heavily on manual interpretation of clinical data, imaging, and laboratory results, are often slow, subjective, and prone to inter-observer variability [2]. This has catalyzed a paradigm shift toward Smart Healthcare Systems (SHS)—integrated ecosystems combining AI, IoMT sensors, cloud computing, and big data analytics to enable real-time, data-driven clinical decision-making [3].

Artificial intelligence, particularly machine learning and deep learning, has demonstrated diagnostic performance comparable to or exceeding that of expert clinicians in several domains. Esteva et al. demonstrated that a deep CNN trained on 129,450 clinical images achieved dermatologist-level classification of skin cancer, with an area under the curve (AUC) exceeding 0.96 [4]. Similarly, Gulshan et al. reported that a deep learning algorithm for detecting diabetic retinopathy from retinal fundus photographs achieved a sensitivity of 97.5% and specificity of

93.4% on validation datasets exceeding 11,000 images [5]. These results illustrate the transformative diagnostic potential of AI when trained on sufficiently large and well-annotated datasets.

Beyond individual diagnosis, AI is increasingly deployed at the population level for public health management. During the COVID-19 pandemic, AI-driven epidemiological models were used extensively for outbreak forecasting, resource allocation, and contact tracing [6]. The pandemic accelerated global investment in digital health infrastructure; according to a 2023 market analysis, the global AI in healthcare market was valued at approximately USD 20.9 billion in 2024 and is projected to grow at a compound annual growth rate (CAGR) of 38.5% through 2030 [7].

This article makes the following contributions:

1. A structured review of AI/ML/DL architectures applied to disease prediction and diagnosis across four major disease categories.
2. A comparative performance analysis (accuracy, sensitivity, specificity, F1-score) drawn from benchmark studies, presented in tabular format.
3. A methodological framework for smart healthcare system design incorporating IoMT, federated learning, and cloud-edge computing.
4. A critical discussion of implementation challenges including data bias, interpretability, and regulatory constraints, along with proposed future directions.

The remainder of this article is organized as follows: Section II presents the literature review; Section III describes the methodology and system architecture; Section IV presents results and comparative analysis; Section V discusses implications and challenges; and Section VI concludes the article.

2. LITERATURE REVIEW

A. AI in Cardiovascular Disease Prediction

Cardiovascular disease (CVD) remains the leading cause of mortality globally, responsible for an estimated 17.9 million deaths annually [1]. Weng et al. evaluated four ML algorithms—random forest, logistic regression, gradient boosting, and neural networks—on a cohort of 378,256 patient records from the UK, reporting that neural network models improved CVD risk prediction accuracy by up to 7.6% compared to the established ACC/AHA risk score [8]. Attia et al. demonstrated that a CNN applied to standard 12-lead ECG data could detect asymptomatic left ventricular dysfunction with an AUC of 0.93, sensitivity of 86.3%, and specificity of 85.7% [9].

B. AI in Diabetes Management

For diabetes, AI models have been applied both for risk stratification and complication prediction. Kavakiotis et al. conducted a systematic review of 40 ML studies related to diabetes prediction, finding that support vector machines (SVM) and ensemble methods consistently outperformed traditional logistic regression, with reported accuracies ranging from 80% to 98% depending on dataset size and feature selection [10]. Continuous glucose monitoring (CGM) devices integrated with LSTM-based forecasting models have enabled prediction of hypoglycemic events up to 30–60 minutes in advance with root mean square errors (RMSE) below 20 mg/dL in several studies [11].

C. AI in Cancer Diagnosis

Deep learning has achieved notable success in oncology imaging. McKinney et al. reported that an AI system for breast cancer screening from mammograms reduced false positives by 5.7% and false negatives by 9.4% compared to radiologists in a UK dataset, and by 1.2% and 2.7% respectively in a US dataset [12]. Coudray et al. showed that CNNs could classify lung cancer histopathology slides into adenocarcinoma versus squamous cell carcinoma with an AUC of 0.97 [13].

D. AI for Public Health Surveillance and Epidemic Forecasting

At the population level, AI-driven epidemiological models gained significant traction during the COVID-19 pandemic. Jewell et al. compared several compartmental and ML-based forecasting models for COVID-19 case prediction, noting that ensemble approaches combining SEIR (Susceptible-Exposed-Infected-Recovered) models with ML calibration reduced forecasting error by 15–20% compared to standalone mechanistic models [14]. Google and Apple's mobility data, combined with AI-based network models, was used to estimate transmission dynamics and inform lockdown policy decisions across over 130 countries [6]. BlueDot, an AI-based epidemic intelligence platform, reportedly flagged the COVID-19 outbreak in Wuhan on 31 December 2019—several days before official WHO notification—by analyzing natural language processing (NLP) outputs from news reports, airline ticketing data, and animal disease networks [15].

E. IoMT and Wearable-Based Continuous Monitoring

The integration of wearable sensors with AI has enabled continuous physiological monitoring outside clinical settings. Perez et al.'s Apple Heart Study, involving 419,297 participants, found that AI-based irregular pulse notifications from smartwatch photoplethysmography had a positive predictive value of 84% for atrial fibrillation detection when confirmed via ECG patch [16]. Such large-scale, real-world deployments highlight the feasibility of population-scale passive health monitoring.

F. Federated Learning for Privacy-Preserving Diagnostics

A key barrier to AI adoption in healthcare is data privacy under regulations such as HIPAA and GDPR. Federated learning (FL) addresses this by training models across decentralized institutional datasets without transferring raw patient data. Rieke et al. demonstrated that FL-trained models for brain tumor segmentation achieved performance within 1–2% of centrally trained models across 10 institutions, while preserving data locality [17]. Sheller et al. similarly showed FL enabled multi-institutional tumor segmentation models to match centrally-trained model performance (Dice score 0.85 vs. 0.86) [18].

Table I. Summary of Key AI Applications in Disease-Specific Diagnosis

Disease Domain	AI Technique	Dataset Size	Reported Accuracy/AUC	Reference
Skin Cancer	Deep CNN (Inception v3)	129,450 images	AUC 0.96	[4]
Diabetic Retinopathy	Deep CNN	11,711 images	Sens. 97.5%, Spec. 93.4%	[5]
Cardiovascular Risk	Neural Network	378,256 records	+7.6% vs. ACC/AHA score	[8]
LV Dysfunction (ECG)	CNN	44,959 patients	AUC 0.93	[9]
Diabetes Risk	SVM/Ensemble	Varies (40 studies)	80–98%	[10]
Breast Cancer	Deep CNN	25,856 (UK) + 3,097 (US)	FP↓5.7%, FN↓9.4% (UK)	[12]
Lung Cancer Histopathology	CNN	1,634 slides	AUC 0.97	[13]
Atrial Fibrillation	ML + PPG	419,297 participants	PPV 84%	[16]

3. METHODOLOGY

A. System Architecture Overview

The proposed smart healthcare system architecture is structured across four functional layers, consistent with established IoMT-AI frameworks [3], [19]:

1. **Perception Layer**: Wearable sensors, IoMT devices, and clinical instruments (ECG, glucometers, pulse oximeters, imaging systems) collect raw physiological and diagnostic data.

2. **Network Layer**: Data transmission via 5G, Bluetooth Low Energy (BLE), and edge gateways, ensuring low-latency communication for time-critical alerts.
3. **Data Processing and AI Layer**: Cloud/edge-based ML/DL pipelines perform preprocessing, feature extraction, model inference, and anomaly detection.
4. **Application Layer**: Clinician dashboards, patient-facing mobile applications, and public health surveillance interfaces present actionable outputs.

B. Data Preprocessing Pipeline

Given the heterogeneity of clinical data (structured EHR fields, unstructured clinical notes, time-series sensor data, and medical images), a multi-modal preprocessing pipeline is essential:

- **Missing data handling**: Multiple imputation by chained equations (MICE) or k-nearest neighbor (KNN) imputation, shown to outperform mean imputation by 12–18% in RMSE reduction on EHR datasets [20].
- **Normalization**: Min-max scaling or z-score standardization applied to continuous physiological variables (heart rate, blood glucose, SpO2).
- **Feature selection**: Recursive feature elimination (RFE) and mutual information filtering to reduce dimensionality prior to model training, improving both training efficiency and interpretability.
- **Class imbalance correction**: Synthetic Minority Oversampling Technique (SMOTE), commonly applied given the rarity of positive disease cases in screening datasets (e.g., disease prevalence often <10% in population screening cohorts) [21].

C. Model Selection Criteria

Model architecture selection depends on data modality:

Data Modality	**Preferred Architecture**	**Justification**
Medical Imaging (X-ray, MRI, histopathology)	CNN (ResNet, Inception, EfficientNet)	Spatial feature extraction
Time-series (ECG, CGM, vitals)	LSTM/GRU, Temporal CNN	Sequential dependency modeling
Structured EHR (tabular)	XGBoost, Random Forest	High performance on tabular data, interpretable feature importance
Clinical text/NLP surveillance	Transformer-based (BERT, BioBERT)	Contextual language understanding
Epidemic forecasting	Hybrid SEIR + ML calibration	Combines mechanistic and data-driven strengths

D. Evaluation Metrics

Standard evaluation employs a confusion-matrix-derived metric suite:

- Accuracy = (TP + TN) / (TP + TN + FP + FN)
- Sensitivity (Recall) = TP / (TP + FN)
- Specificity = TN / (TN + FP)
- F1-Score = 2 × (Precision × Recall) / (Precision + Recall)
- AUC-ROC for threshold-independent discriminative performance

Given the class imbalance typical in disease screening, AUC-ROC and F1-score are prioritized over raw accuracy, as accuracy alone can be misleading when disease prevalence is low [21].

E. Federated Learning Protocol

For multi-institutional deployment, a federated averaging (FedAvg) protocol is adopted [17]:

1. A global model is initialized at a central coordinating server.
2. The model is distributed to participating institutions (hospitals/clinics).
3. Each institution trains the model locally on its private dataset for a fixed number of epochs.
4. Only model weight updates (gradients), not raw data, are transmitted back to the server.
5. The server aggregates updates via weighted averaging based on local dataset size.
6. Steps 2–5 repeat until convergence.

This protocol preserves compliance with data protection regulations while enabling model generalization across diverse patient populations.

4. RESULTS AND COMPARATIVE ANALYSIS

A. Diagnostic Performance Comparison

Aggregating benchmark results reported across the reviewed literature, deep learning models demonstrate consistent superiority over classical ML methods in image-based diagnostic tasks, while ensemble tree-based methods remain competitive for structured/tabular clinical data.

Table II. Comparative Model Performance Across Disease Domains

Task	Best Model	Accuracy	Sensitivity	Specificity	AUC
Skin cancer classification [4]	Inception v3 CNN	—	—	—	0.96
Diabetic retinopathy detection [5]	Deep CNN	—	97.5%	93.4%	0.99
CVD risk prediction [8]	Neural Network	~78.5%	—	—	—
LV dysfunction (ECG) [9]	CNN	—	86.3%	85.7%	0.93
Breast cancer screening [12]	Deep CNN	—	↑ vs. radiologist	↑ vs. radiologist	—
Lung cancer histopathology [13]	CNN (Inception v3)	—	—	—	0.97
Atrial fibrillation detection [16]	ML + PPG	—	—	—	PPV 84%
COVID-19 forecasting (ensemble) [14]	SEIR + ML hybrid	15–20% error reduction vs. mechanistic-only	—	—	—

B. Public Health Surveillance Case Study: COVID-19

During the COVID-19 pandemic, AI-based surveillance platforms were deployed at scale. BlueDot's NLP-based system processed data from over 100,000 sources in 65 languages to detect early outbreak signals [15]. Google/Apple mobility-informed transmission models were adopted by public health agencies in over 130 countries to calibrate non-pharmaceutical intervention timing [6]. Ensemble forecasting models—which combined mechanistic epidemiological models with ML-based calibration layers—reduced short-term case forecasting error by 15–20% relative to standalone SEIR models, according to comparative evaluation across multiple US states [14].

C. Wearable-Based Continuous Monitoring Outcomes

The Apple Heart Study represents one of the largest wearable-AI health studies conducted, enrolling 419,297 participants over 8 months. Of participants who received an irregular pulse notification, 34% subsequently wore an ECG patch, and of those, atrial fibrillation was confirmed in 34% of cases, yielding an overall positive predictive value of 84% when notifications were followed by ECG-patch-confirmed episodes [16]. This demonstrates the feasibility of population-scale passive AI-driven screening, though it also highlights the need for confirmatory clinical-grade testing to manage false-positive burden.

D. Federated Learning Performance Retention

Multi-institutional federated learning studies show minimal performance degradation compared to centralized training. Sheller et al. reported a Dice similarity coefficient of 0.850 for FL-trained brain tumor segmentation models versus 0.862 for centrally-trained models across four institutions—a relative performance gap of under 1.5% [18]. This supports FL as a viable privacy-preserving alternative for cross-institutional AI model development.

5. DISCUSSION

A. Clinical and Public Health Implications

The evidence synthesized in this article demonstrates that AI-driven diagnostic systems can match or exceed expert-level performance in specific, well-defined tasks—particularly image-based classification (dermatology, radiology, ophthalmology, histopathology). However, performance in these studies is typically evaluated on curated, retrospective datasets; prospective, real-world clinical trial validation remains comparatively limited. McKinney et al.'s breast cancer study is among the few evaluated with independent US and UK test cohorts, underscoring the importance of geographic and demographic generalizability testing before clinical deployment [12].

At the public health level, AI-augmented epidemic forecasting demonstrated measurable value during COVID-19, improving short-term forecasting accuracy and enabling earlier outbreak detection through NLP-based surveillance [14], [15]. However, the pandemic also revealed limitations: many early COVID-19 diagnostic AI models developed in 2020 were later found to have high risk of bias due to small, non-representative training cohorts, a concern documented extensively in systematic reviews of COVID-19 prediction models [22].

B. Data Bias and Algorithmic Fairness

A recurring concern across the literature is algorithmic bias arising from non-representative training data. Obermeyer et al. demonstrated that a widely used commercial risk-prediction algorithm, applied to over 200 million patients in the US healthcare system, exhibited significant racial bias—at a given risk score, Black patients were considerably sicker than White patients, because the algorithm used healthcare cost as a proxy for health need, and unequal access to care meant less money was historically spent on Black patients [23]. Correcting this bias increased the percentage of Black patients identified for extra care from 17.7% to 46.5%. This finding illustrates that technically accurate models can nonetheless produce inequitable outcomes if proxy variables encode systemic disparities.

C. Interpretability and Explainable AI (XAI)

Deep learning models, while often highly accurate, function largely as "black boxes," posing challenges for clinical trust and regulatory approval. Techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) have been proposed to improve interpretability by attributing prediction contributions to individual input features [24]. Grad-CAM visualization has similarly been used in medical imaging to highlight regions of an image most influential to a CNN's classification decision, aiding clinician verification of model reasoning [25]. Regulatory bodies, including the FDA, have emphasized the need for "predetermined change control plans" and transparency documentation for AI/ML-based Software as a Medical Device (SaMD) [26].

D. Data Privacy and Regulatory Compliance

Federated learning and differential privacy techniques offer promising pathways to reconcile the data-hungry nature of deep learning with stringent privacy regulations such as HIPAA (US) and GDPR (EU). However, FL introduces its own challenges, including communication overhead, non-independent and identically distributed (non-IID) data across institutions, and vulnerability to model inversion or membership inference attacks, which remain active areas of research [17].

E. Infrastructure and Scalability Challenges

Real-time public health applications require low-latency inference, which has driven interest in edge computing architectures where lightweight model variants (e.g., MobileNet, quantized/pruned models) run directly on IoMT devices rather than relying solely on cloud inference. This reduces latency and bandwidth dependency, which is particularly valuable in resource-constrained or connectivity-limited settings such as rural clinics in low- and middle-income countries [19].

F. Limitations of Current Research

Several limitations constrain the current evidence base:

1. Many high-performing models are validated only on retrospective datasets, limiting prospective clinical applicability.
2. Dataset diversity remains limited, with underrepresentation of certain demographic groups, geographic regions, and rare disease subtypes.
3. Reporting standards vary considerably across studies, complicating direct cross-study comparison (a concern also raised regarding heterogeneous COVID-19 model reporting) [22].
4. Long-term outcome data (i.e., whether AI-assisted diagnosis improves actual patient outcomes, not just diagnostic accuracy) remains sparse.

G. Future Directions

Based on the identified gaps, future research should prioritize:

- Prospective, multi-site randomized clinical trials evaluating AI-assisted diagnostic workflows against standard-of-care.
- Standardized bias auditing frameworks applied prior to clinical deployment, building on lessons from documented algorithmic bias cases [23].
- Wider adoption of federated and privacy-preserving learning frameworks to enable larger, more diverse multi-institutional training datasets without compromising patient privacy.
- Integration of explainable AI techniques as a default requirement, not an optional add-on, for clinical decision-support systems.
- Development of lightweight, edge-deployable models to extend AI-driven diagnostics to low-resource healthcare settings.
- Closer integration between clinical AI systems and public health surveillance infrastructure to enable seamless escalation from individual diagnosis to population-level outbreak detection.

6. CONCLUSION

AI-driven smart healthcare systems are reshaping disease prediction, diagnosis, and public health management through deep learning-based imaging diagnostics, wearable-integrated continuous monitoring, and AI-augmented epidemic forecasting. Evidence reviewed across cardiovascular disease, diabetes, cancer, and infectious disease surveillance demonstrates that AI models frequently achieve diagnostic performance comparable to, or exceeding, expert clinicians under controlled evaluation conditions. However, translating this performance into equitable, trustworthy, and regulator-approved clinical practice requires addressing persistent challenges in algorithmic bias, model interpretability, data privacy, and prospective clinical validation. Federated learning and explainable AI represent particularly promising directions for reconciling the competing demands of diagnostic performance, patient privacy, and clinical trust. Continued interdisciplinary collaboration between clinicians, data scientists, and public health authorities will be essential to ensure AI-driven healthcare innovations translate into measurable improvements in population health outcomes.

References

1. World Health Organization, "Noncommunicable diseases," WHO Fact Sheets, 2023.
2. E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature Medicine, vol. 25, no. 1, pp. 44–56, 2019.
3. S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The Internet of Things for health care: A comprehensive survey," IEEE Access, vol. 3, pp. 678–708, 2015.
4. A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.
5. V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.
6. N. Oliver et al., "Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle," Science Advances, vol. 6, no. 23, 2020.
7. Grand View Research, "Artificial intelligence in healthcare market size, share & trends analysis report," Market Research Report, 2024.
8. S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?," PLOS ONE, vol. 12, no. 4, e0174944, 2017.
9. Z. I. Attia et al., "Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram," Nature Medicine, vol. 25, no. 1, pp. 70–74, 2019.
10. I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," Computational and Structural Biotechnology Journal, vol. 15, pp. 104–116, 2017.
11. K. Zhu, N. Hamilton, N. Kollias, and D. G. Cook, "Deep learning for hypoglycemia prediction in continuous glucose monitoring data," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 5, pp. 1653–1662, 2021.
12. S. M. McKinney et al., "International evaluation of an AI system for breast cancer screening," Nature, vol. 577, no. 7788, pp. 89–94, 2020.
13. N. Coudray et al., "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," Nature Medicine, vol. 24, no. 10, pp. 1559–1567, 2018.
14. N. P. Jewell, J. A. Lewnard, and B. L. Jewell, "Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections," JAMA, vol. 323, no. 19, pp. 1893–1894, 2020.
15. BlueDot Inc., "How BlueDot's human and AI-powered team spotted coronavirus before global health authorities," Company Case Report, 2020.
16. M. V. Perez et al., "Large-scale assessment of a smartwatch to identify atrial fibrillation," New England Journal of Medicine, vol. 381, pp. 1909–1917, 2019.
17. N. Rieke et al., "The future of digital health with federated learning," npj Digital Medicine, vol. 3, no. 1, p. 119, 2020.
18. M. J. Sheller et al., "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," Scientific Reports, vol. 10, no. 1, p. 12598, 2020.
19. M. Chen, Y. Ma, J. Song, C. Lai, and B. Hu, "Smart clothing: Connecting human with clouds and big data for sustainable health monitoring," Mobile Networks and Applications, vol. 21, no. 5, pp. 825–845, 2016.
20. S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in R," Journal of Statistical Software, vol. 45, no. 3, pp. 1–67, 2011.
21. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.
22. L. Wynants et al., "Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal," BMJ, vol. 369, m1328, 2020.
23. Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," Science, vol. 366, no. 6464, pp. 447–453, 2019.
24. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.
25. R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2017, pp. 618–626.
26. U.S. Food and Drug Administration, "Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan," FDA Report, 2021