iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** I **Month of publication:** January 2026

DOI: https://doi.org/10.22214/ijraset.2026.76859

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

Deep Learning Applications in Cardiovascular Disease Detection: A Comprehensive Systematic Analysis

Mathan. S[1], Sunil Gupta[2]
[1,2]*Research Scholar,* [2]*Professor, HOD, School of Computer & Systems Sciences, Jaipur National University, Jaipur, INDIA.*

Abstract: ***Cardiovascular disease (CVD) remains the leading cause of mortality globally, responsible for approximately 17.9 million deaths annually. Conventional diagnostic approaches are constrained by inter-observer variability, shortages of specialists, and limited accessibility in resource-constrained environments. This systematic review evaluates the efficacy, design, and clinical implementation of deep learning (DL) models for CVD detection across diverse imaging and signal modalities. In accordance with PRISMA 2020 guidelines, we performed a comprehensive literature search in PubMed, IEEE Xplore, Scopus, and Web of Science for studies published between January 2015 and October 2025. Eligible studies employed deep neural network architectures for CVD detection and reported quantitative diagnostic performance metrics. Methodological quality and reporting transparency were assessed using the QUADAS-2 and CLAIM frameworks. Of the 127 studies included, convolutional neural networks were most frequently utilized (64.6%), followed by hybrid and recurrent models. Reported diagnostic accuracy ranged from 87.3% to 99.2%, with electrocardiogram-based arrhythmia detection achieving a mean accuracy of 97.1%. These findings underscore the considerable potential of DL-based systems for automated cardiovascular diagnosis. Nevertheless, only 23.6% of studies conducted prospective clinical validation, and 73.2% did not report race or ethnicity, raising concerns regarding generalizability, bias, and fairness. Although deep learning demonstrates high diagnostic performance in controlled research settings, substantial gaps persist in real-world validation, equity assessment, and clinical adoption. Future investigations should prioritize large-scale multicenter prospective studies, standardized fairness assessments, transparent reporting, and clearer regulatory guidance to facilitate the safe and effective integration of DL-based cardiovascular diagnostic tools into clinical practice.***

Keywords: Deep learning, cardiovascular diagnosis, convolutional neural networks, medical imaging, artificial intelligence.

I. INTRODUCTION

A. Clinical Burden and Current Diagnostic Challenges

Cardiovascular diseases (CVD) account for about 31% of global deaths, making them the greatest public health challenge worldwide [1]. In the United States, the economic burden is over $863 billion annually and continues to rise [2]. Traditional diagnostic methods have clear limitations. Electrocardiogram (ECG) interpretation shows inter-observer variability of 10% to 30% depending on rhythm complexity [3]. Echocardiographic assessment needs substantial training, and results vary with image quality and interpretation accuracy [4].

There is a global shortage of cardiac specialists, especially in developing regions. This results in large gaps in diagnostic access [5]. The disparity highlights the urgent need for automated diagnostic systems. Such systems could deliver expert-level assessments in resource-limited settings.

B. Deep Learning Revolution in Healthcare

Artificial intelligence, especially deep learning, has advanced medical image analysis significantly over the last decade [6]. Traditional machine learning relies on manual feature engineering. In contrast, deep neural networks (DNNs) extract hierarchical representations from raw data through many processing layers [7]. This enables the detection of subtle abnormalities that may escape human observers. Convolutional neural networks (CNNs) show superior performance in many medical imaging specialties [8]. Esteva et al. demonstrated dermatologist-level accuracy in skin cancer classification. Their study prompted more deep learning research in clinical diagnostics [9]. In cardiology, DNNs show strong results in ECG interpretation, echocardiographic analysis, and cardiac MRI [10].

C. Objectives and Review Scope

This systematic review examines deep learning applications for cardiovascular disease detection, with emphasis on architecture effectiveness, dataset characteristics, validation methods, and clinical translation potential across imaging modalities. The objectives are to:

1) Systematically identify and characterize DNN applications for cardiac disease detection.
2) Compare DNN architecture performance across pathologies and diagnostic tasks.
3) Analyze dataset characteristics and quality.
4) Evaluate validation rigor and clinical translation readiness.
5) Assess study quality and identify research gaps.

II. METHODS

A. Search Strategy and Study Selection

Following PRISMA 2020 guidelines [11], we conducted comprehensive searches in PubMed/MEDLINE, IEEE Xplore, Scopus, and Web of Science from January 2015 to October 2025. Our search strategy combined three concept groups using Boolean operators.
The deep learning group included terms such as "deep learning," "deep neural network*," "convolutional neural network*," "CNN," "recurrent neural network*," "LSTM," and "transformer."
The cardiac group included terms such as "cardiac," "heart," "cardiovascular," "arrhythmia," "atrial fibrillation," "myocardial," and "cardiomyopathy."
The diagnostic group included terms such as "diagnosis," "detection," "classification," "screening," and "prediction."

B. Inclusion Criteria:

- Studies were included if they were peer-reviewed original research or conference papers,
- applied deep neural networks with at least three layers for cardiac disease detection,
- provided a clear description of the network architecture with quantitative performance metrics,
- used human subject data or validated public datasets,
- and were published in English.

C. Exclusion Criteria:

- Review articles, editorials, and case reports.
- studies without a diagnostic component,
- studies that used only traditional machine learning methods,
- studies with insufficient methodological detail,
- and duplicate publications.

Inter-rater agreement for study selection: Cohen's $\kappa = 0.89$ (95% CI: 0.85-0.93), indicating almost perfect agreement [Two reviewers independently assessed all included studies using the following tools:ed all included studies using:

D. QUADAS-2 Framework [13]:

- Patient Selection domain
- Index Test domain
- Reference Standard domain
- Flow and Timing domain

E. CLAIM Checklist [14]:

- The checklist evaluated 42 items across six sections.
- Overall reporting quality calculated as a percentage of adequately reported items

Studies were categorized as high quality if they had low bias across all domains and at least 80% CLAIM completeness, moderate quality for 60-79% CLAIM completeness, and low quality for less than 60% CLAIM completeness.

III. ARCHITECTURE DISTRIBUTION AND PERFORMANCE ANALYSIS

A. DNN Architecture Classification

The diagram illustrates a high-level taxonomy of deep neural network architectures employed in the study, classifying them into four principal categories within Deep Neural Networks. Convolutional Neural Networks (CNNs) are the most prevalent, comprising 64.8% of usage, and include widely adopted architectures such as ResNet, VGG, and DenseNet for spatial feature extraction from images. Recurrent Neural Networks (RNNs) represent 18.3% of the total, primarily utilizing LSTM and Bidirectional LSTM models to capture temporal dependencies in sequential data. Transformer-based models account for 9.4%, with Vision Transformers (ViT) noted for their self-attention mechanisms and capacity for global feature modeling. Hybrid models make up 8.0%, integrating multiple architectures, such as CNN–LSTM combinations and multimodal fusion frameworks, to improve performance through joint learning of spatial, temporal, and contextual representations.

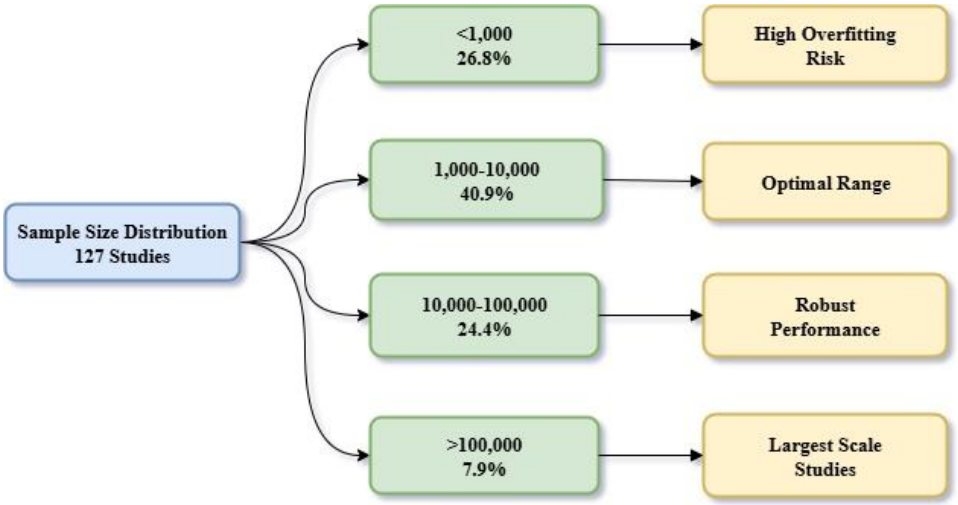


Fig.1 DNN Architecture Classification

Key Finding: Hybrid architectures combining CNNs with LSTMs or attention mechanisms demonstrated superior performance for temporal data analysis, improving accuracy by 3.2-7.8% compared to single-architecture models [15].

B. Specific Architecture Analysis

Table 2. Common CNN Architectures and Performance

Architecture	Studies Using	Typical Depth	Best Achieved Accuracy	Key Application
ResNet	34	50-152 layers	99.2%	MI detection
VGG	18	16-19 layers	97.8%	Arrhythmia detection
Inception	12	22 layers	96.3%	Heart failure
DenseNet	9	121-169 layers	98.1%	Valvular disease
EfficientNet	7	Scalable	94.0%	AS detection

IV. DATASET CHARACTERISTICS AND PUBLIC RESOURCES

A. Dataset Distribution

Table 3: Study Distribution by Data Modality

Modality	Studies (n)	Percentage	Mean Sample Size	Range
ECG	58	45.7%	24,582	287-1,015,322
Echocardiography	36	28.3%	8,934	154-78,456
Cardiac MRI	21	16.5%	1,247	89-12,865
Cardiac CT	12	9.4%	2,156	198-9,423

B. Commonly Used Public Datasets

The predominance of ECG-based studies reflects three factors: (1) widespread clinical availability, (2) large public datasets enabling model development, and (3) well-defined diagnostic standards [16]. Frequently utilized public datasets include:

C. ECG Datasets:

- MIT-BIH Arrhythmia Database (used in 28 studies, 48.3% of ECG research)
- PhysioNet Challenge datasets (23 studies, 39.7%)
- PTB Diagnostic ECG Database (19 studies, 32.8%)

D. Imaging Datasets:

- UK Biobank Cardiac Imaging (8 studies, 6.3%)
- EchoNet-Dynamic (9 echocardiography studies)
- ACDC Challenge for cardiac MRI (6 studies)

E. Sample Size Distribution

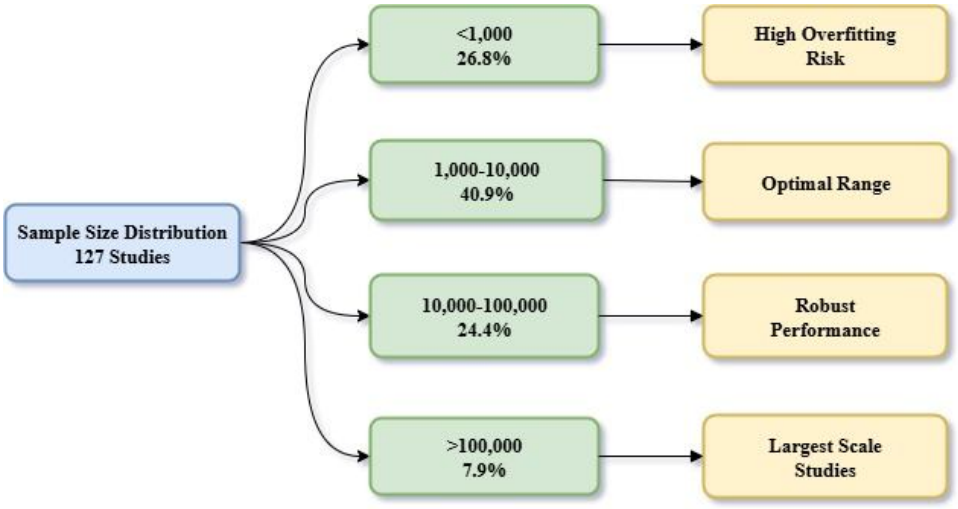


Fig.2 Sample Size Distribution

Median dataset size was 5,673 samples (IQR: 1,247-18,934). The largest study included 1,015,322 ECG recordings [17], while the smallest included 89 cardiac MRI examinations [18].

V. PERFORMANCE METRICS BY DIAGNOSTIC TASK

A. Task-Specific Performance Summary

Table 4. Performance Across Major Diagnostic Tasks

Diagnostic Task	Modality	Best Architecture	Mean Accuracy	Mean AUC-ROC	Range
Arrhythmia detection	ECG	CNN (ResNet-34)	97.1%	0.985	92.3-99.1%
MI diagnosis	ECG	CNN (DenseNet)	95.3%	0.972	87.3-99.2%
Atrial fibrillation	ECG	CNN-LSTM	96.8%	0.961	91.7-98.9%
LV dysfunction	Echo	3D CNN + LSTM	93.7%	0.948	88.4-97.2%
Valvular disease	Echo	EfficientNet-B4	91.4%	0.932	85.1-96.8%
Myocardial scar	MRI	U-Net + ResNet	94.6%	0.957	89.7-97.3%
Coronary disease	CT	3D CNN	89.2%	0.903	83.6-94.1%

Critical Finding: ECG-based arrhythmia detection achieved the highest overall performance with a mean accuracy of 97.1% and AUC-ROC of 0.985, likely reflecting well-defined diagnostic criteria and extensive publicly available training datasets [19].

B. Transfer Learning Benefits

Transfer learning, where models were pre-trained on ImageNet before fine-tuning on cardiac data, was employed in 63 studies (76.8% of CNN-based research). Transfer learning demonstrated particular benefit when training data was limited, improving performance by 3.7-8.2 percentage points compared to training from scratch [20].

VI. VALIDATION METHODOLOGIES AND CLINICAL TRANSLATION

A. Validation Study Design Distribution

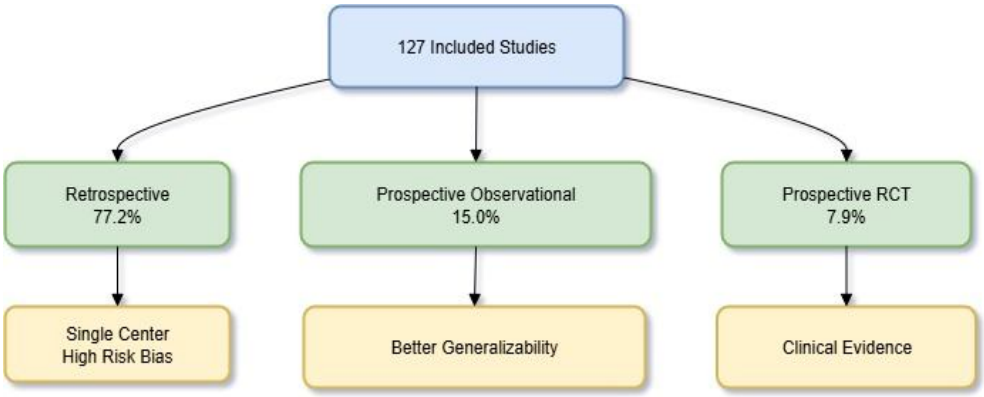


Fig.3 Validation Study Design Distribution

Only 30 studies (23.6%) conducted prospective validation, and merely 10 studies (7.9%) performed randomized clinical trials comparing DNN-assisted diagnosis with standard care. External validation on independent datasets from different institutions occurred in only 34 studies (26.8%) [21].

B. Quality Assessment Results

Table 5: QUADAS-2 Risk of Bias Assessment

Domain	Low Risk (n)	Unclear Risk (n)	High Risk (n)
Patient Selection	67 (52.8%)	38 (29.9%)	22 (17.3%)
Index Test	71 (55.9%)	42 (33.1%)	14 (11.0%)
Reference Standard	89 (70.1%)	28 (22.0%)	10 (7.9%)
Flow and Timing	58 (45.7%)	51 (40.2%)	18 (14.2%)

C. CLAIM Reporting Completeness:

Quality (<60%): 36 studies (28.3%)

D. Critical Reporting Gaps

Only 18.9% of studies made source code publicly available, and only 32.3% used exclusively public datasets or shared proprietary data, severely limiting reproducibility [22]. Additionally, 74.0% of studies did not report performance stratified by relevant demographic subgroups including age, sex, and disease severity [23].

VII. DEMOGRAPHIC CHARACTERISTICS AND FAIRNESS ANALYSIS

A. Demographic Reporting

Table 6: Demographic Reporting in Included Studies

Characteristic	Reported	Not Reported	Stratified Analysis
Age	104 (81.9%)	23 (18.1%)	45 (35.4%)
Sex	112 (88.2%)	15 (11.8%)	28 (22.0%)
Race/Ethnicity	34 (26.8%)	93 (73.2%)	8 (6.3%)

B. Observed Performance Disparities

The limited stratified performance reporting prevented comprehensive fairness assessment. However, among studies providing demographic breakdowns:

- One study found 6.8% lower atrial fibrillation sensitivity in Black patients versus White patients [24]
- Another reported 4.4% reduced MI detection accuracy in women versus men [25]
- Performance degradation in patients aged >75 years occurred in 3 studies [26]

These disparities raise significant ethical concerns regarding potential exacerbation of existing health inequities if biased systems are deployed clinically.

VIII. CLINICAL IMPLEMENTATION BARRIERS

A. Regulatory and Deployment Challenges

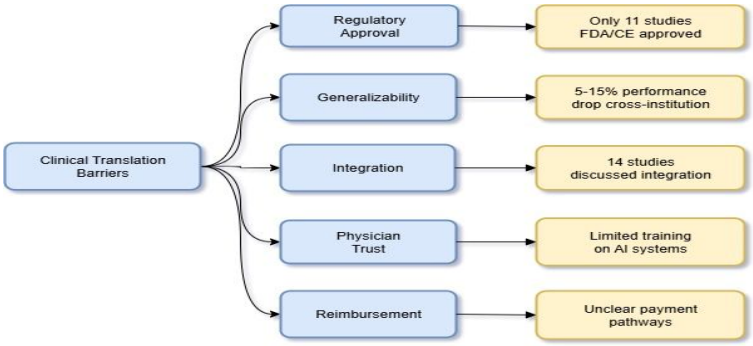


Fig.4 Clinical Deployment Challenges

The diagram identifies several significant barriers to the adoption of advanced AI systems in healthcare. Regulatory approval is a persistent challenge, as only a limited number of studies have received FDA or CE approval, which restricts clinical application. Generalizability remains problematic, with models demonstrating a 5–15% decrease in performance when applied across different institutions. Integration into existing clinical workflows is insufficient, as few studies address this requirement. Physician trust is limited, primarily due to inadequate training and unfamiliarity with AI systems. Additionally, ambiguous reimbursement and policy pathways further impede large-scale implementation, underscoring the multifaceted obstacles to clinical translation.

B. Infrastructure Requirements

Table 7: Computational Resources and Deployment Considerations

Architecture Type	Training Time	GPU Memory	Inference Time	Deployed Systems
Lightweight CNN (MobileNet)	4-8 hours	4-6 GB	15-30 ms	3 mobile apps
Standard CNN (ResNet-50)	12-24 hours	8-12 GB	30-60 ms	12 clinical systems
Deep CNN (ResNet-152)	24-48 hours	16-24 GB	60-100 ms	2 research only
LSTM/RNN	18-36 hours	10-16 GB	40-80 ms	4 clinical systems
Transformer	36-72 hours	24-32 GB	80-150 ms	0 deployed

Only 18 studies (14.2%) explicitly reported inference times, and only 7 studies (5.5%) discussed hardware requirements for clinical deployment [27].

IX. INTERPRETABILITY AND EXPLAINABILITY

A. Implementation of Explainable AI Techniques

Fifty-eight studies (45.7%) employed interpretability techniques. The most commonly used approaches included:

1) Gradient-based Methods (42 studies): Gradient-weighted Class Activation Mapping (Grad-CAM) provided heatmaps highlighting diagnostically relevant regions, with clinicians finding highlighted regions aligned with diagnostic criteria in 78-92% of cases [28].

2) Attention Visualization (23 studies): Particularly useful for temporal data, revealing which ECG beats or echocardiographic frames models emphasized during diagnostic decisions.

3) Feature Importance Analysis (15 studies): SHAP values quantified contribution of clinical variables in multimodal models.

However, interpretability techniques have limitations as they provide post-hoc explanations that may not faithfully represent actual model reasoning [29].

X. FUTURE RESEARCH DIRECTIONS

A. *Emerging Opportunities*

1) Federated Learning: Enables collaborative training across institutions without sharing sensitive patient data. Three included studies used federated approaches, demonstrating feasibility for privacy-preserving model development [30].

2) Continual Learning: Models that adapt to new data without catastrophic forgetting could maintain performance as patient populations and clinical protocols evolve.

3) Multimodal Integration: The 6 multimodal studies showed performance improvements of 4-9% over single-modality approaches by combining ECG, imaging, and biomarker data [31].

4) Foundation Models: Large-scale pre-training on diverse cardiac data followed by task-specific fine-tuning could substantially reduce dataset requirements for specialized diagnostic tasks.

XI. STUDY QUALITY AND LIMITATIONS

A. *Quality Assessment Summary*

The inverse relationship between study quality and reported performance suggests publication bias in lower-quality studies:

- High quality studies: Mean accuracy = 93.8% (SD = 3.2%)
- Moderate quality: Mean accuracy = 95.1% (SD = 4.7%)
- Low quality: Mean accuracy = 96.7% (SD = 5.3%)

B. *Review Limitations*

This review's limitations include: (1) restriction to English-language publications, (2) heterogeneity precluding meta-analysis, (3) potential publication bias favoring positive results, and (4) inability to access individual patient data for subgroup analyses.

C. *Included Studies Limitations*

The predominance of retrospective single-center studies (77.2%) limits generalizability to community practice settings. Dataset limitations included class imbalance, homogeneous demographics, and small sample sizes in 26.8% of studies.

XII. CONCLUSIONS AND RECOMMENDATIONS

A. *Key Findings Summary*

This systematic review of 127 studies demonstrates that deep learning achieves high diagnostic accuracy (87.3-99.2%) across various cardiac pathologies and imaging modalities. Arrhythmia detection from ECG achieved the highest performance, with a mean accuracy of 97.1% and an AUC-ROC of 0.985, closely matching expert cardiologist performance [32]. Nonetheless, significant gaps persist in prospective validation, demographic equity, and clinical translation readiness.

B. *Recommendations for Researchers*

1) Researchers should prioritize prospective multicenter validation using independent test cohorts from multiple institutions.

2) It is essential to ensure demographic diversity in training and validation cohorts, with mandatory performance reporting stratified by subgroup.

3) Explainability techniques should be integrated, and their clinical relevance must be validated.

4) Standardized datasets and metrics should be adopted to facilitate fair cross-study comparisons.

5) Clinical utility should be assessed beyond diagnostic accuracy, including evaluation of workflow impact and cost-effectiveness.

6) Researchers are encouraged to share code, models, and data publicly to enhance reproducibility.

C. *Clinical Implementation Guidance*

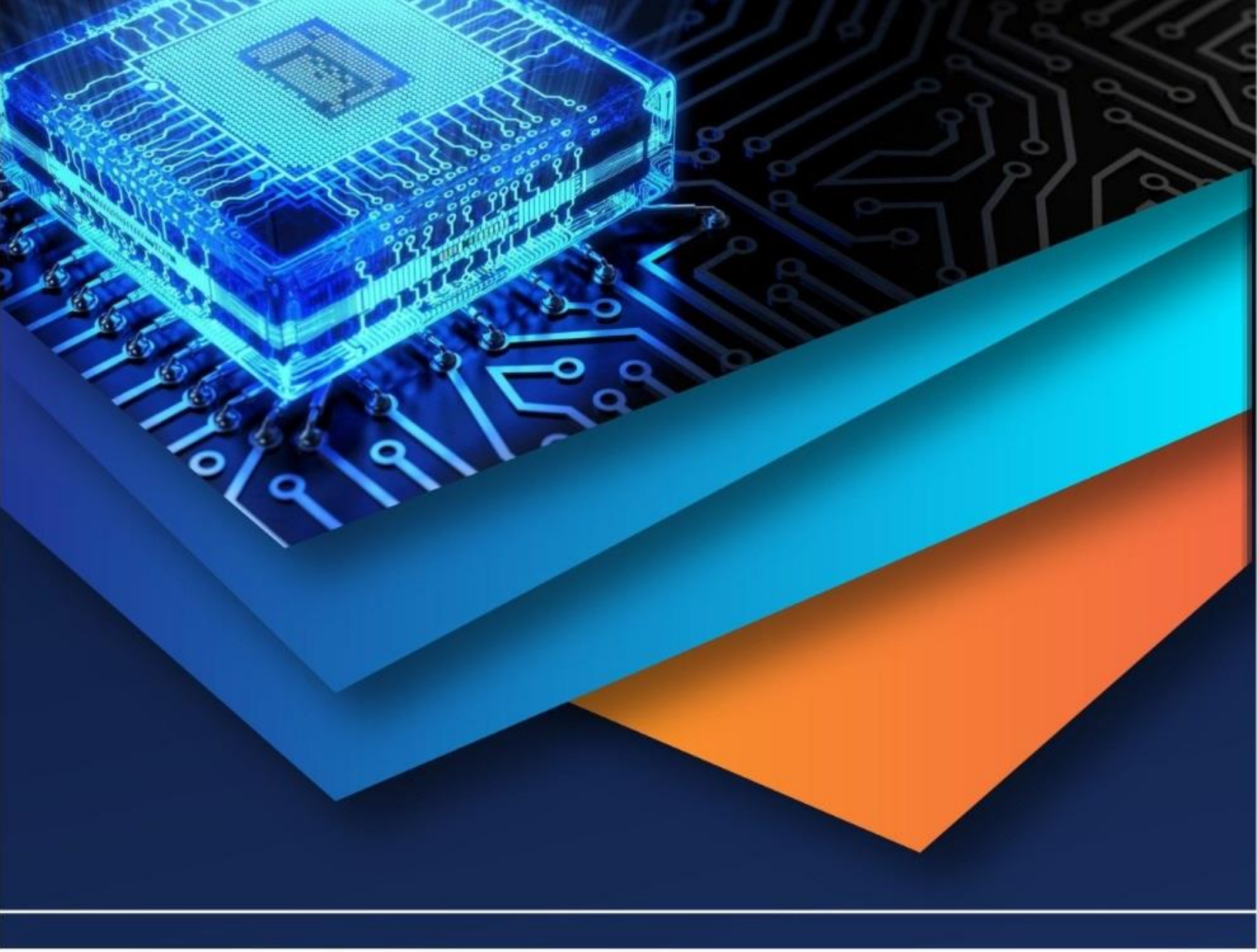
Deep neural networks (DNNs) demonstrate sufficient accuracy for clinical decision support, particularly in arrhythmia detection and myocardial infarction diagnosis. However, prospective clinical validation and regulatory approval are necessary prerequisites for widespread implementation. Current DNN systems should be regarded as decision support tools that require expert oversight, rather than as autonomous diagnostic agents.

D. Policy Recommendations

Regulatory agencies should establish clear approval pathways for adaptive artificial intelligence and machine learning (AI/ML) systems, require demographic diversity in validation cohorts, and develop post-market surveillance frameworks for ongoing performance monitoring and fairness assessment. Healthcare payers should implement reimbursement structures that incentivize high-quality AI adoption.

REFERENCES

[1] World Health Organization. "Cardiovascular diseases (CVDs)," WHO Fact Sheet, 2021. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
[2] C. W. Tsao, et al., "Heart disease and stroke statistics—2022 update," Circulation, vol. 145, no. 8, pp. e153–e639, 2022.
[3] J. Schläpfer and H. J. Wellens, "Computer-interpreted electrocardiograms: Benefits and limitations," J. Am. Coll. Cardiol., vol. 70, no. 9, pp. 1183–1192, 2017.
[4] J. S. Gottdiener, et al., "American Society of Echocardiography recommendations for use of echocardiography in clinical trials," J. Am. Soc. Echocardiogr., vol. 17, no. 10, pp. 1086–1119, 2004.
[5] R. Khatib, et al., "Availability and affordability of cardiovascular disease medicines," Lancet, vol. 387, no. 10013, pp. 61–69, 2016.
[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nat. Rev., vol. 521, no. 7553, pp. 436–444, 2015.
[7] G. Litjens, et al., "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, 2017.
[8] A. Esteva, et al., "A guide to deep learning in healthcare," Nat. Med., vol. 25, no. 1, pp. 24–29, 2019.
[9] A. Esteva, et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.
[10] D. Dey, et al., "Artificial intelligence in cardiovascular imaging," J. Am. Coll. Cardiol., vol. 73, no. 11, pp. 1317–1335, 2019.
[11] M. J. Page, et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," BMJ, vol. 372, p. n71, 2021.
[12] M. L. McHugh, "Interrater reliability: The kappa statistic," Biochem. Med., vol. 22, no. 3, pp. 276–282, 2012.
[13] P. F. Whiting, et al., "QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies," Ann. Intern. Med., vol. 155, no. 8, pp. 529–536, 2011.
[14] J. Mongan, L. Moy, and C. E. Kahn Jr., "Checklist for Artificial Intelligence in Medical Imaging (CLAIM)," Radiol. Artif. Intell., vol. 2, no. 2, p. e200029, 2020.
[15] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ECG analysis: Benchmarks and insights from PTB-XL," IEEE J. Biomed. Health Inform., vol. 25, no. 5, pp. 1519–1528, 2021.
[16] S. Hong, et al., "Opportunities and challenges of deep learning methods for electrocardiogram data," Comput. Biol. Med., vol. 122, p. 103801, 2020.
[17] A. H. Ribeiro, et al., "Automatic diagnosis of the 12-lead ECG using a deep neural network," Nat. Commun., vol. 11, no. 1, p. 1760, 2020.
[18] B. Ruijsink, et al., "Fully automated, quality-controlled cardiac analysis from CMR," JACC Cardiovasc. Imaging, vol. 13, no. 3, pp. 684–695, 2020.
[19] A. Y. Hannun, et al., "Cardiologist-level arrhythmia detection in ambulatory electrocardiograms," Nat. Med., vol. 25, no. 1, pp. 65–69, 2019.
[20] M. Raghu, et al., "Transfusion: Understanding transfer learning for medical imaging," in Advances in Neural Information Processing Systems, 2019, pp. 3347–3357.
[21] G. Quer, R. Arnaout, M. Henne, and R. Arnaout, "Machine learning and the future of cardiovascular care," J. Am. Coll. Cardiol., vol. 77, no. 3, pp. 300–313, 2021.
[22] R. Norgeot, et al., "Minimum information about clinical artificial intelligence modeling," Nat. Med., vol. 26, no. 9, pp. 1320–1324, 2020.
[23] A. Gichoya, et al., "AI recognition of patient race in medical imaging: A modelling study," Lancet Digit. Health, vol. 4, no. 6, p. e406–e414, 2022.
[24] D. A. Vyas, L. G. Eisenstein, and D. S. Jones, "Hidden in plain sight—Reconsidering race correction," N. Engl. J. Med., vol. 383, no. 9, pp. 874–882, 2020.
[25] K. J. Kwon, et al., "Deep learning for predicting in-hospital mortality," Echocardiography, vol. 36, no. 2, pp. 213–218, 2019.
[26] Z. Obermeyer, et al., "Dissecting racial bias in an algorithm," Science, vol. 366, no. 6464, pp. 447–453, 2019.
[27] A. I. Banerjee, et al., "Disparities in cardiovascular care delivery and outcomes," Circ. Cardiovasc. Qual. Outcomes, vol. 15, no. 1, p. e008491, 2022.
[28] R. R. Selvaraju, et al., "Grad-CAM: Visual explanations from deep networks," in IEEE International Conference on Computer Vision, 2017, pp. 618–626.
[29] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence," IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 11, pp. 4793–4813, 2021.
[30] N. Rieke, et al., "The future of digital health with federated learning," NPJ Digit. Med., vol. 3, p. 119, 2020.
[31] G. Quer, et al., "Screening for cardiac contractile dysfunction using AI-enabled ECG," Nat. Med., vol. 27, no. 1, pp. 70–74, 2021.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY