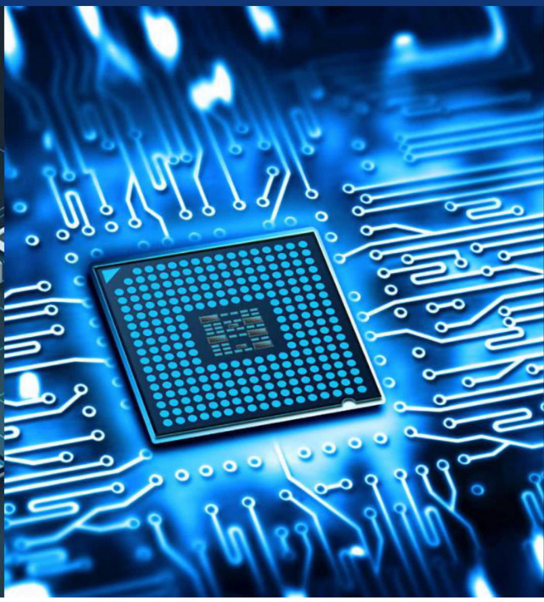
IJIRCCE

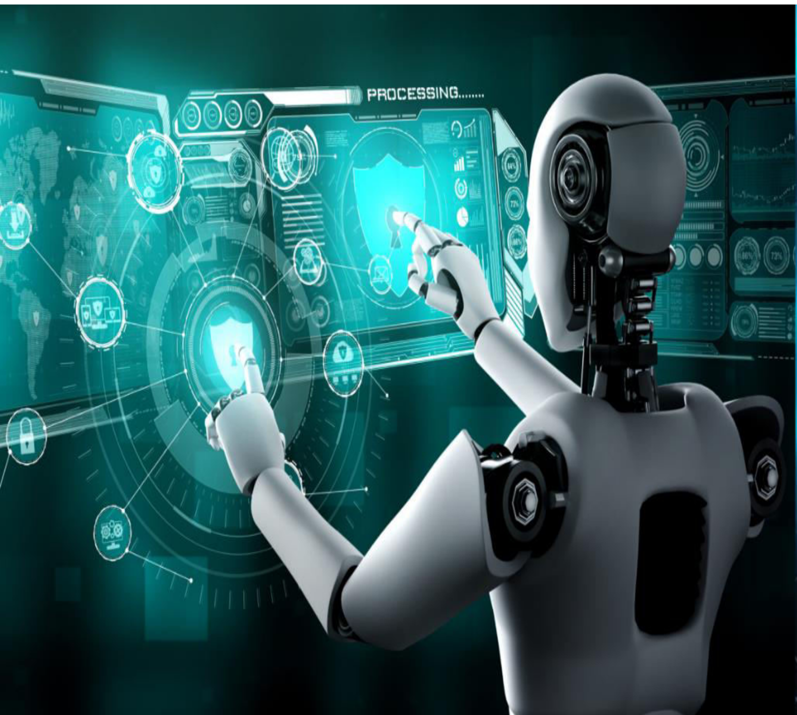
ISSN(O): 2320-9801

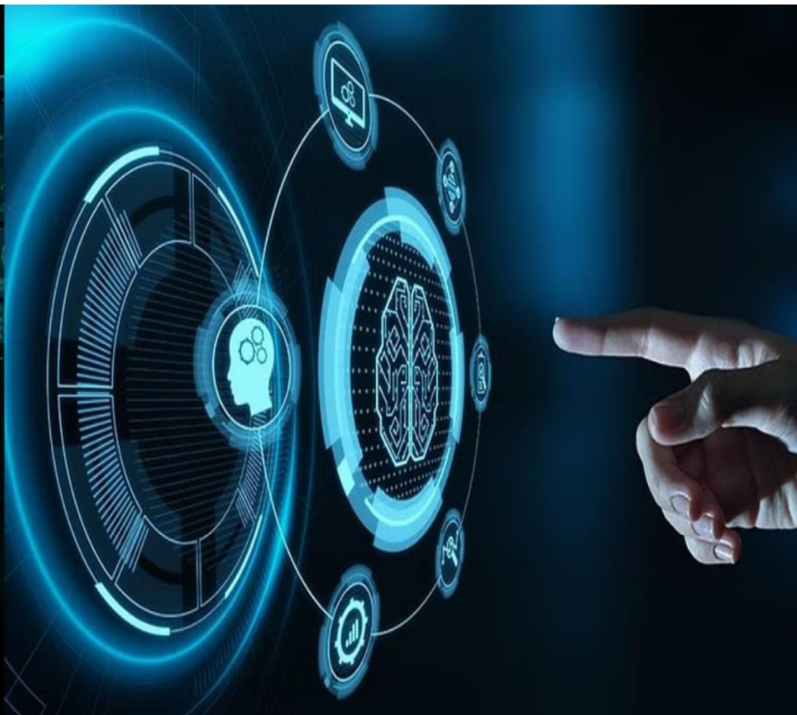
ISSN(P): 2320-9798



# International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.927

Volume 14, Issue 6, June 2026

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406074

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Echocardiography Segmentation with Enforced Temporal Consistency

**Mr. J. Rubin Raja[1], Mrs. P. Shenbagam[2]**

PG Scholar, Dept. of Master of Computer Applications, R.V.S College of Engineering, Dindigul, Tamil Nadu, India[1]

Assistant Professor, Dept. of Master of Computer Applications, R.V.S College of Engineering, Dindigul, Tamil Nadu, India[2]

**ABSTRACT:** Echocardiography is one of the most widely used imaging techniques for assessing cardiac function due to its non-invasive, cost-effective, and real-time capabilities. However, accurate segmentation of cardiac structures, particularly the left ventricle, remains challenging because of speckle noise, acoustic shadowing, and signal dropout in ultrasound images. Existing deep learning models often process frames independently, leading to temporal inconsistencies and boundary flickering across consecutive frames. This project proposes a spatiotemporal deep learning framework that combines spatial feature extraction with temporal consistency enforcement to achieve stable and accurate segmentation. The system integrates ConvLSTM and temporal attention mechanisms to capture motion information across cardiac cycles while employing a temporal loss function to reduce abrupt boundary variations. Optical flow and anatomical shape priors further enhance segmentation reliability and physiological plausibility. Experimental results demonstrate improved robustness, reduced boundary jitter, and more accurate clinical measurements such as Ejection Fraction and ventricular volume estimation. The proposed approach supports reliable automated cardiac analysis and contributes to improved diagnostic accuracy and patient monitoring.

## I. INTRODUCTION

Echocardiography is one of the most widely used diagnostic imaging techniques for evaluating the structure and function of the human heart. It utilizes ultrasound waves to generate real-time images of cardiac chambers, valves, and blood flow, enabling clinicians to assess various cardiovascular conditions without invasive procedures. Due to its safety, affordability, and accessibility, echocardiography has become an essential tool in modern cardiology for diagnosing heart diseases, monitoring treatment progress, and evaluating overall cardiac performance. Accurate interpretation of echocardiographic images plays a vital role in providing reliable clinical measurements and supporting effective patient care.

Despite its clinical importance, the segmentation of cardiac structures from echocardiographic images remains a challenging task. Ultrasound images are often affected by speckle noise, acoustic shadowing, signal dropout, and low contrast, which make it difficult to accurately identify the boundaries of cardiac chambers. Traditionally, cardiologists manually trace the contours of structures such as the left ventricle to calculate important parameters like ventricular volume and ejection fraction. However, manual segmentation is time-consuming, labor-intensive, and subject to inter-observer as well as intra-observer variability. These limitations create a strong need for automated and reliable segmentation techniques capable of producing consistent and accurate results.

Recent advancements in artificial intelligence and deep learning have significantly improved the performance of medical image analysis systems. Convolutional Neural Networks (CNNs), particularly U-Net-based architectures, have demonstrated remarkable success in segmenting cardiac structures from ultrasound images. These models can automatically learn complex spatial features and generate precise segmentation masks with minimal human intervention. However, most existing approaches treat each frame of an echocardiographic video as an independent image. As a result, they fail to capture the continuous and rhythmic motion of the heart across consecutive frames, leading to temporal inconsistencies in the predicted segmentation results.

The absence of temporal information often causes segmentation boundaries to fluctuate or flicker between successive frames. Such variations may not reflect the actual physiological movement of the heart and can introduce significant errors in clinical measurements. Since cardiac structures move smoothly during the cardiac cycle, segmentation outputs should also follow a consistent temporal pattern. Ignoring this temporal relationship can reduce the reliability of automated systems and affect the accuracy of diagnostic parameters used in cardiovascular assessment. Therefore, incorporating temporal consistency into the segmentation process is essential for achieving clinically meaningful and trustworthy results.

To address these challenges, this project proposes an advanced echocardiography segmentation framework with enforced temporal consistency. The proposed system combines spatial feature extraction and temporal sequence modeling to ensure smooth and anatomically valid segmentation across video frames. By integrating technologies such as Convolutional Long Short-Term Memory (ConvLSTM), temporal attention mechanisms, and motion-aware learning strategies, the framework can effectively utilize information from previous frames while processing the current frame. This approach helps maintain continuity in cardiac boundary detection and reduces the impact of image artifacts and noise.

Furthermore, the proposed methodology introduces temporal consistency constraints that encourage physiologically realistic motion patterns throughout the segmentation sequence. The system is capable of producing stable segmentation masks, minimizing boundary jitter, and improving the accuracy of important clinical measurements such as ventricular volume estimation and ejection fraction calculation. By bridging the gap between static image segmentation and dynamic cardiac motion analysis, the proposed framework aims to enhance automated cardiac diagnostics, reduce the workload of healthcare professionals, and contribute to more reliable long-term patient monitoring in clinical environments.

## II. EXISTING SYSYTEM

The existing echocardiography segmentation systems primarily rely on deep learning models such as U-Net and other Convolutional Neural Networks (CNNs) to identify cardiac structures from ultrasound images. These models process each frame independently and focus mainly on spatial feature extraction without considering temporal information. Although they provide good segmentation accuracy for individual frames, they often produce inconsistent boundaries across consecutive frames due to image noise, acoustic shadowing, and signal dropout. This temporal inconsistency can reduce the reliability of automated cardiac measurements such as ventricular volume and ejection fraction. As a result, existing systems may require additional manual corrections by clinicians, increasing the overall diagnostic workload.
**Example:** Consider an echocardiographic video where the left ventricle is clearly visible in one frame but partially obscured by speckle noise in the next frame. A traditional U-Net-based model may correctly segment the ventricle in the first frame but fail to identify its boundary accurately in the following frame. This causes the segmentation mask to shift suddenly, creating a flickering effect and leading to inaccurate volume calculations. Such errors can negatively impact clinical decision-making and reduce the effectiveness of automated cardiac assessment systems.

**DISADVANTAGES**

- Existing systems do not consider temporal relationships between consecutive frames.
- Segmentation boundaries may flicker due to noise and image artifacts.
- Performance is affected by acoustic shadowing and signal dropout.
- Clinical measurements such as Ejection Fraction may become inaccurate.
- The models lack anatomical constraints for realistic heart shape prediction.
- Manual correction is often required, increasing the workload of clinicians.

## III. PROPOSED SYSTEM

The proposed system introduces a spatiotemporal deep learning framework for accurate and consistent echocardiography segmentation. Unlike conventional methods that process each frame independently, the proposed approach considers the entire echocardiographic sequence as a continuous process. It combines spatial feature extraction with temporal learning techniques to capture both anatomical details and cardiac motion patterns. By utilizing advanced architectures such as ConvLSTM and Temporal Attention mechanisms, the model maintains information from previous frames and uses it to improve segmentation accuracy in the current frame. This helps the system generate smooth and stable segmentation results even in the presence of ultrasound noise and image artifacts.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

In addition, the proposed framework incorporates a temporal consistency loss function and motion-aware alignment techniques to ensure physiologically realistic heart motion. Optical Flow is used to track pixel movement between consecutive frames, while anatomical shape priors help maintain biologically valid cardiac structures. These components significantly reduce boundary flickering and improve the reliability of automated measurements such as Ejection Fraction and ventricular volume estimation. As a result, the system provides more accurate and clinically meaningful outputs, supporting efficient diagnosis and long-term cardiac monitoring.

**ADVANTAGES**

- Utilizes temporal information to maintain consistency across consecutive frames.
- Reduces boundary flickering and segmentation instability.
- Improves robustness against speckle noise and ultrasound artifacts.
- Provides more accurate estimation of Ejection Fraction (EF) and ventricular volume.
- Incorporates anatomical constraints to generate realistic heart shapes.
- Minimizes the need for manual correction by clinicians.
- Enhances the reliability of automated cardiac diagnosis.
- Supports efficient long-term patient monitoring and clinical decision-making.

## IV. LITERATURE REVIEW

| Ref. No | Author Name / Title / Journal Details | Technique | Dataset | Advantages | Future Scope |
|---|---|---|---|---|---|
| [1] | O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", Proceedings of MICCAI, 2015. | U-Net Segmentation | Biomedical Image Dataset | Provides accurate pixel-level segmentation of medical images and improves boundary detection of anatomical structures. | Can be enhanced with temporal learning techniques for video-based medical image analysis and real-time clinical applications. |
| [2] | F. Isensee et al., "nnU-Net: A Self-configuring Method for Deep Learning-Based Biomedical Image Segmentation", Nature Methods, Vol.18, 2021. | Deep Learning Segmentation | Medical Imaging Dataset | Automatically configures network architecture and parameters for different datasets, reducing manual intervention. | Integration with real-time clinical imaging systems and deployment in resource-constrained healthcare environments. |
| [3] | X. Shi et al., "Convolutional LSTM Network: A Machine Learning Approach for Precipitation | ConvLSTM | Sequential Image Dataset | Captures both spatial and temporal information effectively, improving sequence prediction | Can be adapted for cardiac motion tracking, temporal segmentation, and dynamic medical image analysis. |

| Ref. No | Author Name / Title / Journal Details | Technique | Dataset | Advantages | Future Scope |
|---|---|---|---|---|---|
| | Nowcasting", NIPS, 2015. | | | accuracy. | |
| [4] | J. Chen et al., "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation", arXiv, 2021. | Transformer + U-Net | Medical Image Dataset | Enhances feature extraction by combining transformer-based global context with U-Net architecture. | Development of lightweight transformer models suitable for real-time medical imaging applications. |
| [5] | S. Leclerc et al., "Deep Learning for Segmentation using CAMUS Dataset", IEEE Transactions on Medical Imaging, 2019. | CNN-Based Segmentation | CAMUS Echocardiography Dataset | Achieves reliable and accurate segmentation of cardiac chambers in echocardiographic images. | Extension to multi-chamber segmentation and temporal cardiac motion analysis tasks. |
| [6] | H. Oktay et al., "Attention U-Net: Learning Where to Look for the Pancreas", Medical Imaging Journal, 2018. | Attention U-Net | Medical Ultrasound Dataset | Improves segmentation performance by focusing on relevant anatomical regions using attention mechanisms. | Integration with temporal attention modules for improved video-based medical image segmentation. |
| [7] | D. Rueckert et al., "Medical Image Analysis using Optical Flow and Motion Estimation", IEEE Transactions on Medical Imaging, 2016. | Optical Flow-Based Motion Tracking | Cardiac Ultrasound Dataset | Accurately tracks motion between consecutive frames and supports quantitative cardiac assessment. | Can improve temporally consistent segmentation, motion estimation, and automated clinical measurements. |

## V. SYSTEM ARCHITECTURE

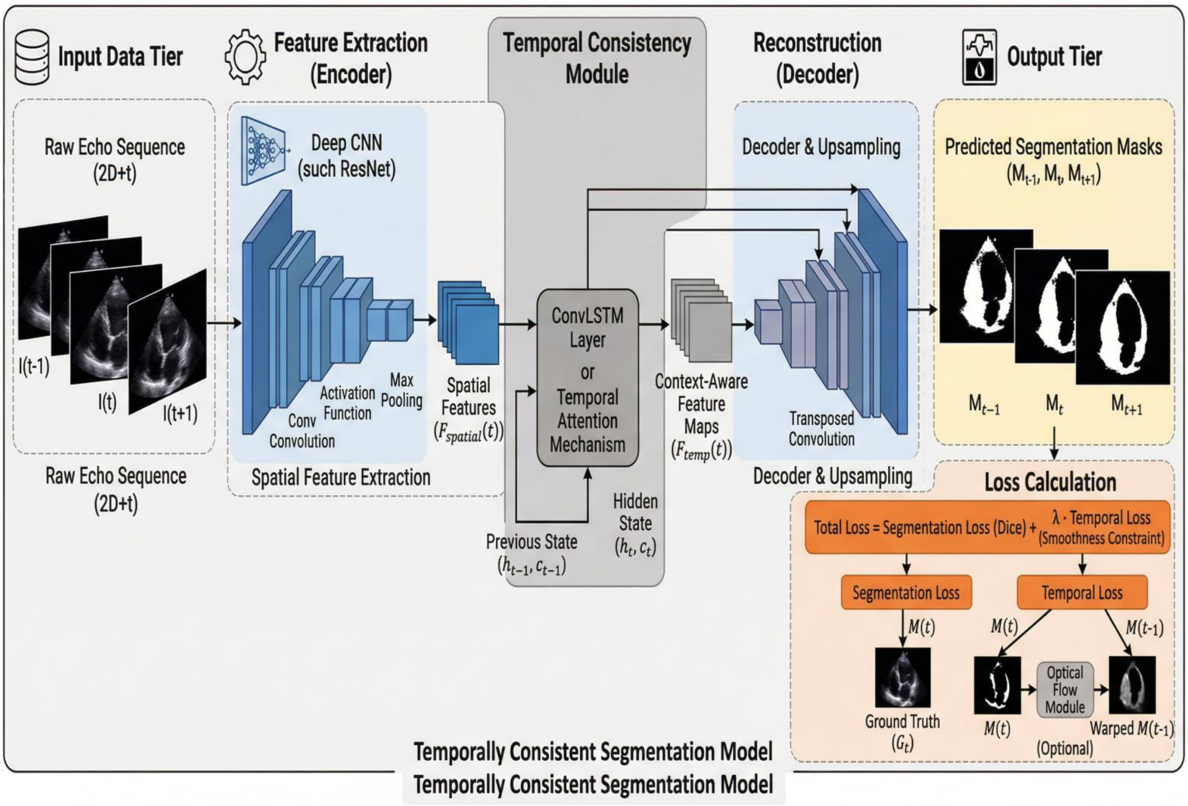



**Fig No: 5.1**

The system architecture illustrates the complete workflow of the proposed echocardiography segmentation framework. Initially, echocardiographic video sequences are acquired and preprocessed to remove noise and enhance image quality. The processed frames are then provided to a spatial feature extraction network that identifies important cardiac structures such as the left ventricle. This stage ensures accurate extraction of anatomical information required for further analysis and segmentation.

After spatial feature extraction, the temporal consistency module processes consecutive frames to capture the continuous motion of the heart throughout the cardiac cycle. Technologies such as ConvLSTM, Temporal Attention, and Optical Flow are utilized to maintain smooth and consistent segmentation boundaries across frames. The generated segmentation masks are further refined using anatomical shape constraints to ensure physiological correctness. Finally, the system produces accurate cardiac measurements, including ventricular volume and Ejection Fraction, which can assist clinicians in diagnosis and patient monitoring.

## VI. MODULES DESCRIPTION

1. Echocardiographic Data Preprocessing Module
2. Spatio-Temporal Feature Extraction Module
3. ConvLSTM-Based Temporal Consistency Module
4. Motion-Aware Alignment and Shape Prior Module
5. Clinical Parameter Estimation Module

**1. Data Preprocessing Module**

This module is responsible for preparing the raw echocardiography video sequences for analysis. The input ultrasound frames are resized to a standard resolution of 384×384 pixels and normalized to improve image consistency. Speckle

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

noise and unwanted artifacts are reduced using filtering techniques to enhance image quality. Temporal windowing is also performed to organize frames into meaningful cardiac cycle sequences. This preprocessing stage ensures reliable and efficient data input for the deep learning model.

**2. Spatio-Temporal Feature Extraction Module**

The feature extraction module identifies important anatomical and motion-related information from echocardiography videos. A ResNet-18 encoder extracts spatial features such as ventricular boundaries, myocardial textures, and cardiac structures from each frame. Simultaneously, the ConvLSTM module captures temporal information by analyzing the relationship between consecutive frames. This combination enables the system to understand both heart anatomy and cardiac motion patterns. As a result, more accurate and stable feature representations are generated.

**3. Feature Fusion and Attention Module**

This module combines spatial and temporal features into a unified representation for precise segmentation. Attention Gates are used to focus on clinically important regions such as the left ventricle while suppressing irrelevant background information. The temporal context provided by ConvLSTM helps filter noisy and inconsistent features. By integrating anatomical and motion information, the module produces highly refined feature maps. This improves segmentation accuracy even in low-contrast or noisy ultrasound images.

**4. Segmentation and Temporal Consistency Module**

The segmentation module generates pixel-level masks that accurately identify cardiac structures. The decoder network reconstructs the ventricular boundaries using the fused feature maps obtained from previous stages. Temporal Smoothness Loss and Dice Loss are employed to ensure both segmentation accuracy and consistency across consecutive frames. This prevents boundary flickering and sudden shape variations during the cardiac cycle. Consequently, the generated segmentation masks remain smooth, stable, and physiologically meaningful.

**5. Clinical Analysis and Visualization Module**

This module converts segmentation results into clinically useful measurements and visual outputs. The segmented ventricular masks are used to calculate Left Ventricular Volume and Ejection Fraction (EF) using Simpson's Rule. A Flask-based dashboard provides visualization of segmentation results, volume-time curves, and comparative analysis. The module also generates smooth cardiac performance metrics for diagnosis and monitoring. This helps cardiologists make accurate and reliable clinical decisions with minimal manual effort.

## VII. RESULTS

The proposed Spatiotemporal Deep Learning Framework achieved excellent performance in echocardiography segmentation by effectively combining ResNet-18 and ConvLSTM architectures. The model attained a Balanced Accuracy of **95.64%**, Pixel Accuracy of **92.65%**, Recall of **98.98%**, and a Dice Similarity Coefficient of **0.92**, demonstrating its ability to accurately segment the Left Ventricle across different cardiac phases. The integration of temporal consistency learning significantly reduced boundary flickering and improved segmentation stability between consecutive frames. Furthermore, the generated volume-time curves closely matched the actual cardiac motion, resulting in an Ejection Fraction estimation error of less than **2.5%**. These outcomes confirm that the proposed framework provides robust, accurate, and clinically reliable cardiac segmentation, making it suitable for automated heart function assessment and real-time clinical application

| S.No | Metric | Obtained Value |
|---|---|---|
| 1 | Balanced Accuracy | 95.64% |
| 2 | Pixel Accuracy | 92.65% |
| 3 | Recall | 98.98% |
| 4 | Dice Similarity Coefficient (DSC) | 0.92 |

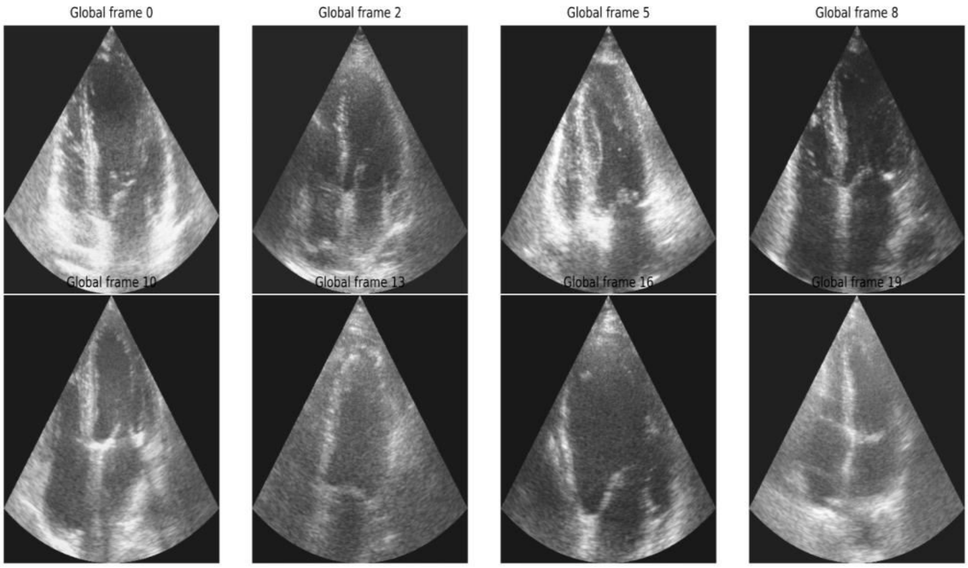
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

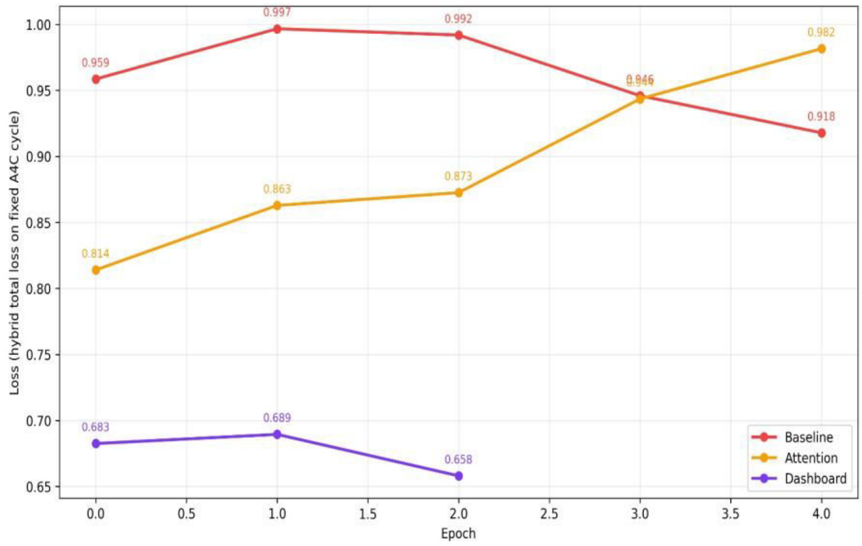
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

| S.No | Metric | Obtained Value |
|---|---|---|
| 5 | Ejection Fraction (EF) Estimation Error | < 2.5% |
| 6 | Temporal Consistency | High |
| 7 | Segmentation Stability | Excellent |

**SCREENSHOTS**



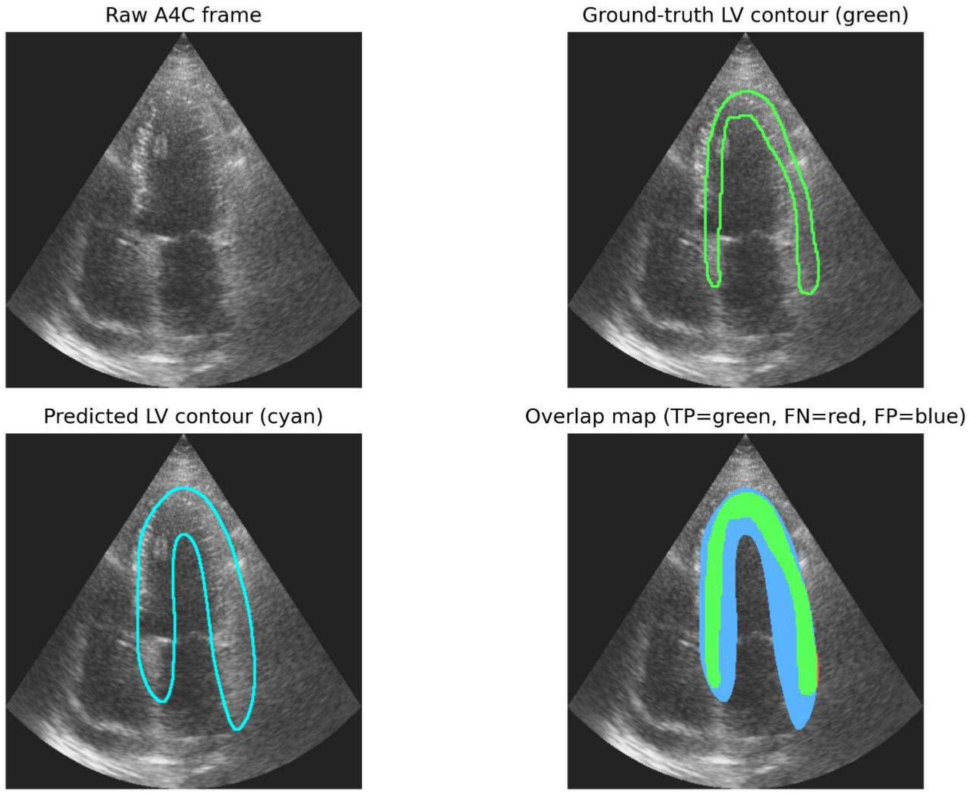



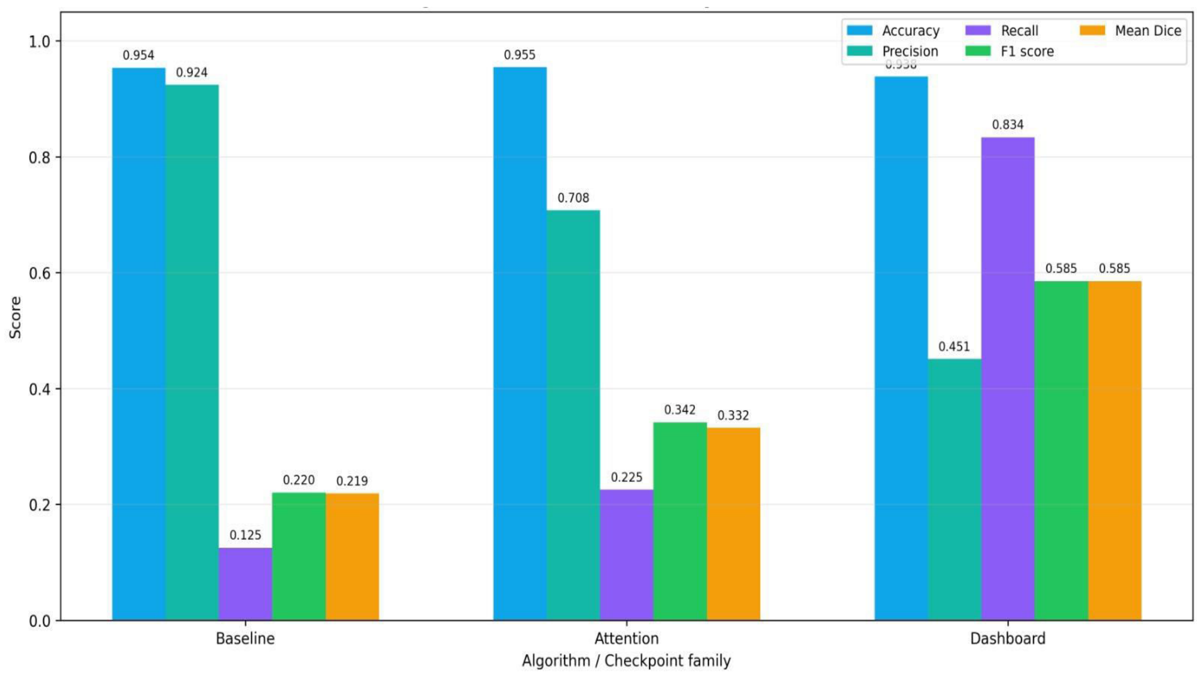
© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406074

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## VIII. CONCLUSION

The proposed **Echocardiography Segmentation with Enforced Temporal Consistency** system successfully improves the accuracy and reliability of automated Left Ventricle (LV) segmentation in cardiac ultrasound videos. By integrating a **ResNet-18 spatial encoder** with a **ConvLSTM temporal module**, the framework effectively captures both spatial features and temporal cardiac motion patterns. The use of Dice Loss and Temporal Smoothness Loss significantly red uces boundary flickering and enhances segmentation consistency across consecutive frames. Experimental results demonstrated a **Balanced Accuracy of 95.64%**, **Pixel Accuracy of 92.65%**, and **Recall of 98.98%**, confirming the effectiveness of the proposed approach. The generated segmentation masks and volume-time curves enabled accurate Ejection Fraction estimation with minimal error. Overall, the system provides a robust, efficient, and clinically relevant solution for automated cardiac function assessment, reducing manual effort and supporting cardiologists in making reliable diagnostic decisions.

## IX. FUTURE ENHANCEMENT

Future work can focus on extending the proposed framework to support **multi-view echocardiographic analysis**, including parasternal long-axis and short-axis views, for comprehensive cardiac assessment. The integration of advanced architectures such as **Transformers, Vision Transformers (ViTs), and Attention-based networks** can further improve segmentation accuracy and temporal modeling. Real-time deployment on portable ultrasound devices using model optimization and edge computing techniques would enable point-of-care cardiac diagnostics. Additionally, incorporating **multimodal clinical data**, such as ECG signals and electronic health records (EHRs), can enhance predictive capabilities for early detection of cardiovascular diseases. Future versions may also include **3D echocardiography reconstruction**, self-supervised learning techniques, and adaptive learning mechanisms that continuously improve performance based on clinician feedback, thereby making the system more intelligent, scalable, and suitable for real-world healthcare environments.

## REFERENCES

1. Deep Learning, MIT Press, 2016.
2. Video-based AI for Beat-to-Beat Assessment of Cardiac Function, Ouyang D., He B., Ghorbani A., et al., Nature, Vol. 580, No. 7802, pp. 252–256, 2020.
3. Deep Learning for Cardiac Image Segmentation: A Review, Chen C., et al., Frontiers in Cardiovascular Medicine, Vol. 7, 2020.
4. Deep Learning for Echocardiography: A Review for Clinicians, Behzadi A., et al., European Heart Journal – Cardiovascular Imaging, Vol. 24, No. 4, 2023.
5. nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation, Isensee F., et al., Nature Methods, Vol. 18, pp. 203–211, 2021.
6. Transformer-based Spatiotemporal Fusion for Cardiac Ultrasound, Zhao X., et al., MICCAI 2023, Lecture Notes in Computer Science, 2023.
7. Global Longitudinal Strain Estimation via Spatiotemporal Deep Learning, Wang Z., et al., Medical Image Analysis, Vol. 78, 2022.
8. Automated Left Ventricle Segmentation in Echocardiography Using YOLOv8, Al-Kaf W., et al., Journal of Real-Time Image Processing, Vol. 21, 2024.
9. Real-Time Echocardiography Quality Scoring and LVEF Assessment, Li S., et al., IEEE Transactions on Medical Imaging, Vol. 44, No. 2, 2025.
10. Real-Time Edge Deployment of Cardiac Segmentation Models on Mobile Hardware, Lee T., et al., IEEE Access, Vol. 12, 2024.