IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# VAE-DRIVEN CARDIAC DISEASE PREDICTION USING MULTIMODAL FUSION

**[1]Ms. Anju G, [2]Abhinaya K, [3]Anamika K, [4]Gopika Pradeep, [5]Sneha Ajith**

[1]Professor, [2]Student, [3]Student, [4]Student, [5]Student
Department of Computer Science and Engineering
St. Thomas College of Engineering and Technology, Kannur, Kerala, India

*Abstract:* **Cardiac diseases, including Severe Left Ventricular Hypertrophy (SLVH), Dilated Left Ventricle (DLV) and Ejection Fraction (EF) abnormalities, contribute to significant morbidity and mortality worldwide. Early detection and accurate diagnosis are crucial for effective management and improved patient outcomes. Traditional diagnostic methods often rely on single modality data, which may overlook the complex nature of cardiac health. This paper proposes an advanced multimodal deep learning framework that integrates echocardiographic images with CXR structured data and CXR imagery to enhance prediction accuracy. By leveraging the complementary strengths of these diagnostic modalities, the system provides a more comprehensive understanding of cardiac health, aiding early detection of SLVH, DLV and EF abnormalities. The approach uses techniques like Variational Auto Encoders (VAEs) for data fusion, EfficientNetB3 for feature extraction, and attention mechanisms like SE-Block and CBAM to focus on clinically relevant features. This frame work addresses the limitations of single-modality analysis and enables timely clinical interventions, ultimately improving the management of cardiac diseases.**

***Index Terms*** **– Cardiac diseases, EfficientNetB3, Variational Auto Encoders, SE-Block, CBAM**

## I. INTRODUCTION

Cardiac diseases represent a major cause of morbidity and mortality globally, affecting the heart's structure and function. These conditions, including Severe Left Ventricular Hypertrophy (SLVH), Dilated Left Ventricle (DLV), and abnormalities in Ejection Fraction (EF), are critical indicators of heart health. SLVH involves thickening of the left ventricle, often due to hypertension or other cardiovascular issues, while DLV is characterized by the dilation of the left ventricle, resulting in a weakened heart muscle. Early detection and accurate diagnosis of these conditions are essential for initiating timely interventions, preventing further complications, and improving patient outcomes.

Recent advancements in medical imaging and machine learning provide an opportunity to enhance cardiac disease detection. Traditional diagnostic methods often rely on single modality data, which may overlook the complexities of a patient's condition. Integrating diverse data sources such as echocardiographic images and chest X-ray (CXR) structured data can offer a more comprehensive understanding of cardiac health. By leveraging the complementary strengths of these modalities, multimodal deep learning approaches can improve diagnostic accuracy, enabling earlier identification of SLVH, DLV, and EF abnormalities.

This paper proposes a multimodal deep learning framework that integrates echocardiographic images with CXR data to enhance the prediction accuracy of cardiac diseases. The system utilizes advanced techniques like Variational Auto Encoders (VAEs) for data fusion, EfficientNetB3 for feature extraction, and attention mechanisms to highlight clinically relevant features. By overcoming the limitations of single modality analysis, this approach provides a more accurate diagnosis, supports timely clinical interventions, and ultimately contributes to the effective management of cardiac diseases.

### *A. PROBLEM STATEMENT*

Severe Left Ventricular Hypertrophy (SLVH) and Dilated Left Ventricle (DLV) are serious cardiac disorders that require timely and accurate diagnosis for effective management. Conventional diagnostic methods often rely on single data sources, such as imaging or clinical evaluations, leading to incomplete assessments and missed diagnoses. While current predictive models excel in either image analysis or textual data interpretation, they struggle to integrate these modalities effectively, which limits their diagnostic power. This highlights the need for innovative approaches that combine various diagnostic tools, improving predictive accuracy and enabling earlier interventions. For instance, integrating data from radiographic imaging and echocardiography with clinical features could create a more comprehensive understanding of the patient's condition. Although deep learning shows promise in cardiac care, a gap remains in integrating radiographic and echocardiographic data for accurate diagnosis.

Addressing these challenges is crucial for advancing cardiac care, improving patient outcomes, and developing models that provide a comprehensive understanding of cardiac health, ultimately allowing clinicians to make more informed decisions and offer personalized treatment plans. By bridging this gap, the medical field can move towards more holistic and proactive approaches to diagnosing and managing complex cardiac conditions.

## II. LITERATURE SURVEY

Recent literature on cardiac disease diagnosis highlights the increasing use of advanced machine learning techniques, particularly deep learning, to enhance predictive accuracy and patient outcomes. Studies show that multimodal learning, which integrates various data sources like echocardiography, MRI, and CT scans, offers a comprehensive view of cardiac health. However, there is a notable gap in effectively combining radiographic and echocardiographic data, with many models excelling in either image analysis or textual data interpretation but struggling to integrate both. This gap underscores the need for innovative approaches that utilize the strengths of both imaging and structured clinical data to improve early detection and diagnosis of cardiac conditions.

### *A. ECG Signals Deep Compressive Sensing Framework Based on Multiscale Feature Fusion and SE Block*

This paper introduces a deep compressive sensing frame work designed specifically for ECG signal reconstruction, incorporating multiscale feature fusion and Squeeze-and Excitation (SE) blocks to enhance signal quality. The approach normalizes ECG signals, utilizes successive convolutional layers for effective compression, and employs Long Short-Term Memory (LSTM) networks to capture temporal dependencies for more accurate reconstruction. Through multiscale feature fusion, the model extracts both local and global patterns, while SE blocks refine feature extraction by emphasizing relevant features. Experimental results show that the proposed method outperforms existing techniques, achieving a percent root difference (PRD) of 1.55% and a signal-to-noise ratio (SNR) of 37.66 dB at a sensing rate of 0.4, demonstrating robust signal preservation even at low sensing rates. This framework is particularly beneficial for real-time ECG monitoring in clinical settings, offering a more comfortable, less invasive patient experience by reducing the number of electrodes needed. Despite its strengths, the framework is limited to ECG signal processing and cannot directly diagnose specific cardiac conditions, focusing solely on improving signal compression and reconstruction.

### *B. DASMcC: Data Augmented SMOTE Multi-Class Classifier for Prediction of Cardiovascular Diseases Using Time Series Features*

This paper introduces a novel approach for predicting cardiovascular diseases (CVD) using a multi-class classifier and data augmentation techniques like SMOTE (Synthetic Minority Over-Sampling Technique) to address class imbalance. The study focuses on analyzing 12-lead ECG data to classify four types of CVD, highlighting the importance of early detection for improving patient outcomes. The methodology includes data preprocessing, feature extraction, and evaluation of five classifiers, with the XG Boost model achieving the highest performance, attaining an accuracy of 93.0% and a recall rate of 90.0%. The framework demonstrates the potential of machine learning in smart healthcare systems for real time CVD prediction and management. However, the study is limited to ECG data, without integrating other data types or exploring fusion techniques that could further enhance predictive accuracy. The proposed approach showcases the potential of using advanced machine learning methods for the early detection of CVD, which could ultimately lead to more timely and effective interventions. However, the reliance solely on ECG data may limit its generalizability, as integrating diverse data sources like patient demographics or lifestyle factors could provide a more holistic view of cardiovascular health.

### *C. VAE-Driven Multimodal Fusion for Early Cardiac Disease Detection*

This paper presents a comprehensive approach to early detection of cardiac diseases, specifically Severe Left Ventricular Hypertrophy (SLVH) and Dilated Left Ventricle (DLV), by leveraging a multimodal deep learning model that integrates diverse data sources, including chest X-ray images and structured clinical records from a substantial dataset of 70,000 medical records from Columbia University. To address data imbalance, the authors employed the Synthetic Minority Over-sampling Technique (SMOTE) to generate additional synthetic records, enhancing the model's learning capacity. The architecture incorporates EfficientNetB3 for image analysis, attention mechanisms like Squeeze-and-Excitation Blocks (SE-Blocks) and Convolutional Block Attention Modules (CBAM) to re fine feature extraction, and Transformer Encoders to process structured data effectively.

Furthermore, the integration of Variational Auto Encoders (VAEs) facilitates the dimensionality reduction of high-dimensional data, allowing for a more nuanced fusion of multimodal features. The model achieved notable accuracy improvements, with increases of 5.43% for SLVH and 14.13% for DLV over existing methods, demonstrating high Area Under the Curve (AUC) values across various disease stages. Overall, this research underscores the potential of advanced AI methodologies in enhancing diagnostic capabilities and improving patient outcomes in cardiac healthcare.

TABLE I. LITERATURE SURVEY TABLE

| PAPER | DESCRIPTION | AUTHOR |
|---|---|---|
| ECG Signals Deep Compressive Sensing Framework Based on Multiscale Feature Fusion and SE Block | This study introduces an efficient ECG compressive sensing method using feature fusion, SE blocks, and LSTM, achieving high-quality signal reconstruction with low error and strong performance. | JING HUA, JIAWEN ZOU, JUE RAO, HUA YIN, JIE CHEN |
| DASMcC: Data Augmented SMOTE Multi-Class Classifier for Prediction of Cardiovascular Diseases Using Time Series Features | This study develops a multi-class ECG-based classifier to detect four cardiovascular diseases using time-domain features and machine learning, showing strong potential for smart healthcare use. | NIDHI SINHA1, M. A. GANESH KUMAR, AMIT M. JOSHI, LINGA REDDY CENKERAMADDI |
| VAE-Driven Multimodal Fusion for Early Cardiac Disease Detection | This study presents a multimodal deep learning model that integrates chest X-ray images and structured CXR data using EfficientNetB3, attention mechanisms, and VAEs to improve early detection of chronic cardiac conditions with superior accuracy and robustness. | JUNXIN WANG , JUANEN LI , RUI WANG, XINQI ZHOU |

## III. METHODOLOGY

The proposed system follows a structured pipeline starting with comprehensive data collection. Echocardiographic images, chest X-rays, and structured clinical data are gathered from a diverse patient population, ensuring inclusivity across demographics and disease stages. Collaborations with healthcare institutions ensure ethical data acquisition while maintaining patient privacy. Longitudinal data is incorporated to analyse disease progression and treatment outcomes over time. The pre-processing phase involves resizing and normalizing medical images to a consistent format, along with applying augmentation and de-noising techniques to enhance image quality. Structured data is standardized, missing values are handled appropriately, and Synthetic Minority Over-sampling Technique (SMOTE) is used to balance class distributions. The dataset is then divided into training, validation and testing subsets.

The model architecture leverages a multimodal fusion approach, integrating features from echocardiograms, CXR images, and structured clinical data. EfficientNetB3 is utilized for robust image feature extraction, enhanced further with attention mechanisms like SE Block and CBAM to focus on critical regions. A hybrid model structure combines convolutional neural networks for image data with fully connected layers for structured data, culminating in a decision tree classifier for interpretable results. Training employs binary cross-entropy loss and incorporates SMOTE for class balancing, with performance evaluated using metrics such as accuracy, AUC, and F1-score, supported by k-fold cross-validation. Learning rate scheduling and early stopping are applied to optimize training and avoid overfitting. External validation is conducted using independent datasets, with clinical expert collaboration to ensure practical relevance and performance across varied demographics. The model is continuously updated with new data to enhance accuracy and adaptability over time. Additionally, the system is designed for real-time deployment, making it suitable for integration into clinical workflows. Its modular architecture allows easy scalability and adaptation to other disease domains in the future.

## IV. PROPOSED SYSTEM

The proposed system uses a multimodal deep learning framework to improve the detection and diagnosis of Severe Left Ventricular Hypertrophy (SLVH), Dilated Left Ventricle (DLV), and Ejection Fraction (EF) abnormalities. By combining echocardiographic images with structured clinical data, it offers a more accurate and comprehensive assessment than traditional single-modality methods. EfficientNetB3, SE Blocks, and CBAM enhance image feature extraction, while VAEs and Transformer encoders ensure effective data fusion and interpretation. A Decision Tree classifier improves classification accuracy, and anomaly detection helps identify rare cardiac conditions. In addition to diagnosis, the system includes a treatment recommendation module that provides personalized therapy plans based on patient data. It supports integration with hospital EHR systems and is designed for real-time use in clinical settings. With cloud-based deployment, it offers scalability for telemedicine and remote care. The model has been validated through extensive ablation studies, showing significant performance gains in terms of accuracy, precision, recall, and F1-score over existing multimodal methods. High AUC values further confirm its robustness in disease progression prediction. Overall, this approach enhances diagnostic precision, supports early detection, facilitates continuous monitoring, and delivers tailored treatment strategies for improved patient outcomes and better resource utilization in healthcare environments.

### *A. Integration of Echo Images*

Echocardiographic images offer real-time insights into the heart's structure and function, helping to visualize chambers, valves, and blood flow. By integrating these images into our deep learning model, we can detect subtle heart abnormalities like SLVH and

DLV more accurately. Attention mechanisms like SE-Block and CBAM help the model focus on key features, improving diagnostic precision. Temporal analysis further supports tracking disease progression over time.

### *B. Ejection Fraction (EF)*

Ejection Fraction measures how much blood the left ventricle pumps with each heartbeat and is essential for diagnosing heart failure. Our system uses echocardiographic and structured clinical data to estimate EF accurately through EfficientNetB3 and Transformer encoders. It helps monitor heart function, detect early declines in EF, and guide timely medical interventions.

### *C. Treatment Recommendation Module*

This module provides personalized treatment suggestions based on detected heart conditions and patient-specific data. It analyzes clinical guidelines and patient history to recommend actions like lifestyle changes, medications, or surgeries. By considering individual health profiles, it supports informed, data-driven decisions and ongoing care adjustments for better patient outcomes.

## V. PROPOSED SYSTEM DESIGN

The architecture shown in Fig 1 illustrates a comprehensive architecture for the proposed model, integrating multiple data modalities to enhance cardiac disease prediction. This architecture is structured to effectively process and analyze echocardiographic images, chest X-ray (CXR) images, and structured clinical data, allowing for a holistic approach to diagnosing cardiac conditions. By leveraging advanced ma chine learning techniques, including EfficientNetB3 for image feature extraction and Transformer encoders for structured data, the system aims to improve diagnostic accuracy and provide valuable insights for clinical decision-making.

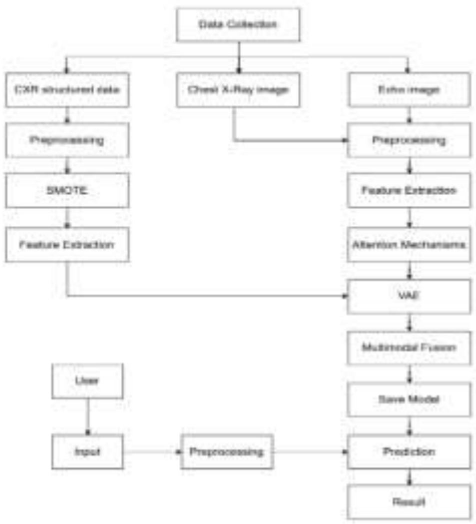


Figure 1. Architecture Diagram

1. Input Layer: Users provide diverse data inputs, including structured CXR data, chest X-ray, and echocardiographic images. This step gathers essential information for analysis, ensuring a comprehensive evaluation.
2. Data Preprocessing: The input data undergoes cleaning and normalization to remove noise and inconsistencies. This improves model performance by ensuring high quality training data.
3. Feature Extraction: EfficientNetB3 extracts features from chest X-ray and echocardiographic images, leveraging its accuracy in image classification. A Transformer encoder processes structured clinical data, dynamically weighing feature importance using self-attention.
4. SMOTE: The Synthetic Minority Over-sampling Technique (SMOTE) balances the dataset by generating synthetic samples, addressing class imbalances to improve model learning.
5. Attention Mechanisms: SE-Block recalibrates channel-wise feature responses, while CBAM applies spatial and channel attention to refine feature maps, improving focus on critical information.
6. Variational Auto Encoders (VAE): VAEs encode data into a latent space, capturing essential representations for better prediction and analysis. This enhances the model’s ability to understand underlying data patterns.
7. Multimodal Fusion: Features from different modalities, including structured data, X-ray, and echocardiographic images, are combined for a more comprehensive representation, improving predictive capabilities.
8. Output Layer: The fused data undergoes analysis to predict cardiac conditions using advanced machine learning techniques. A decision tree classifier enhances interpretability, ensuring accurate and transparent diagnoses

## VI. RESULT

The proposed cardiac condition detection system enhances diagnostic accuracy by integrating EfficientNetB3 for feature extraction and a Decision Tree classifier for classification. It processes echocardiographic images, chest X-ray (CXR) images, and structured clinical data, enabling a comprehensive assessment of Severe Left Ventricular Hypertrophy (SLVH), Dilated Left Ventricle (DLV), and abnormalities in Ejection Fraction. By leveraging attention mechanisms like SE Block and CBAM, the system refines feature representation, improving prediction reliability.

### *A. Admin Web Page*

The Admin plays a key role in managing the platform, overseeing user data and treatment information to ensure accuracy and efficiency. They verify and manage hospital staff accounts, assign user roles, and remove unauthorized users. In treatment management, the Admin can update disease and treatment details—such as for SLVH and DLV—ensuring guidance is current with medical research. They also monitor system performance, troubleshoot issues, enforce security protocols, and maintain data integrity. This oversight supports secure access, accurate diagnostics, and better patient outcomes.


Figure 1. Admin Home Page

### *B. User Web Page*

Users, including doctors and medical staff, manage patient data and support diagnosis through three main features: Personal Information, Patient Data, and Records. They can update their profile details for secure access, input comprehensive patient info including X-ray and echo images, and record vital signs such as blood pressure, cholesterol, glucose levels, and heart rate to aid in diagnosing conditions like SLVH, DLV, and EF abnormalities. The Patient Data section also allows entry of symptoms and medical history, ensuring a complete clinical profile for accurate analysis. The Records section allows viewing past reports, tracking progress over time, and making informed treatment decisions. Search and filter tools help prioritize care by severity, identify high-risk patients, and improve data access efficiency. This streamlined process enhances workflow, supports timely interventions, and ensures better coordination among healthcare teams for improved patient care.


Figure 2. User Home Page

### *C. Prediction*

The Prediction Page allows healthcare professionals to input key patient data and upload chest X-ray and echocardiographic images for automated cardiac condition detection using deep learning. It collects details like vitals, medical history, and structured data to identify SLVH, DLV, and EF abnormalities. Real-time monitoring supports early diagnosis and risk assessment. With dropdowns and formatted fields for accuracy, the system analyzes inputs and delivers AI-based predictions, aiding timely, informed clinical decisions and improving patient outcomes.

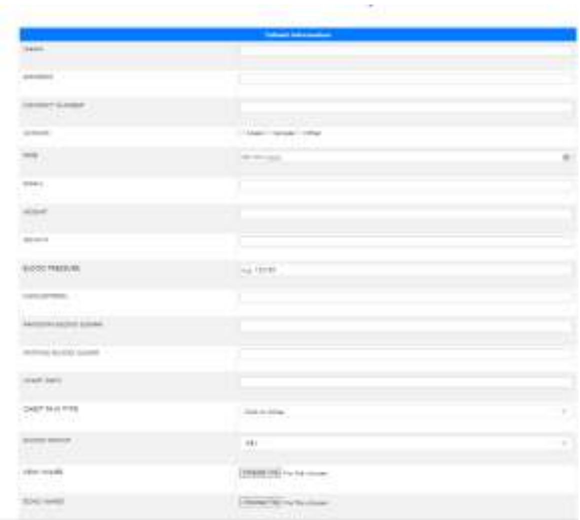

Figure 3. Prediction

***D. Result***

This page lets healthcare professionals access patient diagnostic histories, view test results, track disease progression, and assess treatment effectiveness. A searchable table shows details like date, patient name, ejection fraction, diagnosis (SLVH, DLV, BOTH, NORMAL), treatment, and report view. By monitoring trends, clinicians can detect issues early, adjust care, and personalize treatment. The system supports informed decisions, continuity of care, and efficient, secure record access.

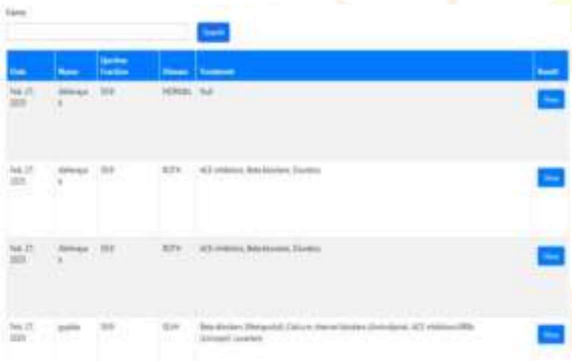

Figure 4. Result

## References

[1] J. Hua, J. Zou, J. Rao, H. Yin, and J. Chen, "ECG signals deep compressive sensing framework based on multiscale feature fusion and SE block," IEEE Access, vol. 11, pp. xxxx-xxxx, Sep. 2023, doi: 10.1109/ACCESS.2023.3316487.

[2] N. Sinha, M. A. Ganesh Kumar, A. M. Joshi, and L. R. Cenkeramaddi, "DASMcC: Data augmented SMOTE multi-class classifier for prediction of cardiovascular diseases using time series features," IEEE Access, vol. 11, pp. xxxx-xxxx, Oct. 2023, doi: 10.1109/ACCESS.2023.3325705.

[3] J. Wang, J. Li, R. Wang, and X. Zhou, "VAE-Driven Multimodal Fusion for Early Cardiac Disease Detection," IEEE Access, vol. 12, pp. 12345 12356, Jun. 2024, doi: 10.1109/ACCESS.2024.3420444.

[4] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for con volutional neural networks," in Proc. 36th Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105–6114. [Online]. Available: http://proceedings. mlr.press/v97/tan19a.html

[5] M. McGilvray, J. Heaton, A. Guo, M. F. Masood, B. P. Cupps, M. Damiano, M. K. Pasque, and R. Foraker, "Electronic health recordbased deep learning prediction of death or severe decompensation in heart failure patients," JACC, Heart Failure, vol. 10, no. 9, pp. 637–647, 2022.

[6] J. Lu, D. Batra, D. Parikh, and S. Lee, "ViLBERT: Pretraining taskagnostic visiolinguistic representations for vision-and-language tasks," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2019.

[7] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7132–7141.

[8] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," in Proc. 38th

Int. Conf. Mach. Learn. (ICML), in Proceedings of Machine Learning Research, vol. 139, 2021, pp. 8748–8763. [Online]. Available: https://proceedings.mlr.press/v139/radford21a.html

[9] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 3–19.

[10] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in Proc. 2nd Int. Conf. Learn. Represent. (ICLR), 2014.

[11] H. M. Shepherd, J. Heaton, T. Marghitu, and L. Bai, "Applying deep learning in heart failure: Hospital readmission is not like other health quality metrics," MedRxiv, Mar. 2024, doi: 10.1101/2024.03.27.