Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 25 May 2026
Revised 26 June 2026
Accepted 21 July 2026

Natural Resources for Human Health 2026, Vol. 6 (9s): 634–644

KEYWORDS:
biomedical signal processing, multimodal deep learning, cross-domain learning, domain adaptation, attention-based fusion, disease prediction

# Robust Cross-Domain Multimodal Deep Learning for Biomedical Signal Interpretation and Disease Prediction

Dr. Suvarna Sunil Nirmal[1], Dr. Sneha Arjun Khaire[2], Rageshri Bakare[3], Anuradha S Deshpande[4*], Dr. Minal Vilas Gade[5], Amit Kurmi[6]

[1] Associate Professor, Sanjivani College of Engineering, Kopargaon, India
[2] Assistant Professor, K. K. Wagh Institute of Engineering Education and Research, Nashik, India
[3] ECE Department, School of Engineering and Sciences, MIT Art, Design and Technology University, India
[4] Associate Professor, Faculty of Science and Technology. JSPM University, Pune, India
[5] Assistant Professor, E&TC Department, Sandip Institute of Technology & Research Centre (SITRC), Nashik, India
[6] Cummins College of Engineering for Women, Nagpur, India

**ABSTRACT:** Biomedical signal analysis is essential for continuous patient monitoring and AI-assisted clinical decision-making using multimodal physiological data. This study proposes a robust cross-domain multimodal deep learning framework that integrates modality-specific encoders, adaptive attention-based fusion, and domain-adversarial learning to improve generalization across heterogeneous biomedical datasets. The framework jointly analyzes electrocardiogram (ECG), photoplethysmography (PPG), and intensive care unit (ICU) physiological signals, including heart rate, blood pressure, respiratory rate, and oxygen saturation, from the publicly available PTB-XL, BIDMC, and MIMIC-IV datasets. Explainability is achieved through Integrated Gradients and modality-level attribution. Experimental results achieved 94.8% accuracy, 93.7% F1-score, and 96.1% ROC-AUC, outperforming CNN, LSTM, Transformer, and tensor fusion baselines. Cross-domain evaluation improved the F1-score by 7–10% over tensor fusion methods while maintaining robustness under dataset shifts, noisy inputs, and missing modalities. These results demonstrate that adaptive multimodal fusion with domain-invariant learning enhances the robustness, interpretability, and clinical applicability of AI-assisted biomedical signal analysis and disease prediction.

## 1. INTRODUCTION

Biomedical signal analysis has become an important component of modern healthcare systems because it enables continuous and non-invasive monitoring of physiological activity. Signals such as electrocardiograms (ECG), electroencephalograms (EEG), and photoplethysmography (PPG) provide complementary information about cardiac and neurological function, making them useful for diagnosis, patient monitoring, and clinical decision support. The widespread adoption of wearable devices and IoMT platforms has accelerated the use of AI for automated biomedical signal interpretation and clinical decision support [1], [2].

Deep learning models including CNNs, recurrent networks, and Transformers have demonstrated strong performance in arrhythmia detection, seizure prediction, and physiological monitoring [3], [4], [5]. Although these approaches have achieved strong results in controlled settings, models trained on a single signal type often fail to represent the complexity of physiological processes. Many clinical conditions involve interactions across multiple systems, which cannot be fully captured using a single

biosignal. This limitation has led to increased interest in multimodal biomedical learning, where multiple physiological signals such as ECG, EEG, and PPG are analyzed together to improve diagnostic consistency. By combining complementary signals, multimodal approaches help reduce ambiguity in predictions and improve clinical reliability [2], [6]. Empirical studies have reported improvements in tasks such as cardiovascular monitoring and physiological state estimation when multiple biosignals are jointly modeled [7], [8].

The effectiveness of multimodal learning largely depends on how information from different signals is combined. Existing approaches are commonly categorized into early fusion, late fusion, and hybrid fusion strategies. Early fusion combines features before model training, allowing direct interaction between modalities but often increasing sensitivity to noise and differences in signal scale. Late fusion trains separate models for each modality and combines their outputs, offering flexibility but limited interaction between signals. Hybrid fusion methods attempt to balance both approaches by combining intermediate representations or using attention-based mechanisms to selectively integrate information across modalities [6], [9]. Recent studies increasingly explore hybrid designs because they provide better adaptability when dealing with heterogeneous or incomplete biomedical data [2], [9].

Despite progress in multimodal modeling, a key challenge remains in ensuring that models perform consistently across different datasets and clinical environments. Biomedical signals collected from different hospitals, devices, and patient populations often follow different distributions, which leads to significant performance drops when models are applied outside their training domain. This issue has driven research in transfer learning, adversarial domain adaptation, and domain generalization. Transfer learning allows knowledge reuse across datasets but assumes some level of similarity between domains [10], [11]. Adversarial domain adaptation reduces distribution gaps using adversarial training but typically requires access to target domain data during training [12], [13]. Domain generalization avoids this requirement by learning representations that are expected to remain stable across unseen environments, making it more suitable for real-world deployment scenarios [14], [15].

Alongside generalization issues, practical deployment also requires models that can handle signal noise, missing channels, and limited interpretability. Biomedical recordings are often affected by motion artifacts, sensor noise, and incomplete measurements, which can reduce model reliability. At the same time, deep learning models are often difficult to interpret, which limits their acceptance in clinical practice. This has created a need for methods that not only integrate multiple modalities and handle domain variation but also provide meaningful explanations for their predictions [2], [16], [17].

Motivated by these challenges, this work focuses on developing a cross-domain multimodal learning framework for biomedical signal interpretation and disease prediction. The proposed approach is designed to integrate multiple physiological signals, learn representations that are less sensitive to dataset-specific variations, handle incomplete or noisy inputs, and provide interpretable outputs suitable for clinical analysis.

Unlike existing multimodal biomedical learning approaches that primarily focus on single-domain learning or fixed fusion strategies, the proposed framework jointly integrates adaptive attention-based multimodal fusion, domain-adversarial representation learning, and explainability within a unified architecture. The contributions of this study can be summarized as follows:

- a cross-domain multimodal learning framework for biomedical signal interpretation and disease prediction,
- an adaptive attention-based fusion mechanism for heterogeneous physiological signals,
- domain-invariant representation learning to improve generalization across datasets,
- integration of explainability components for interpretability of predictions, and
- extensive evaluation under cross-domain, noisy, and missing-modality conditions.

## 2. METHODOLOGY

### *2.1 Overall Framework Architecture*

This study proposes a Robust Cross-Domain Multimodal Deep Learning framework for disease prediction using heterogeneous biomedical signals. The framework integrates ECG, PPG, and ICU-derived physiological time-series data sourced from PTB-XL, BIDMC, and MIMIC-IV PhysioNet datasets. The objective is to learn a unified diagnostic representation that remains invariant to dataset-specific variations while preserving modality-specific physiological information.

Let $X^{ecg} \in R^{\{T \times L\}}$, $X^{ppg} \in R^{\{T \times L\}}$, and $X^{icu} \in R^{\{T \times K\}}$ denote ECG, PPG, and ICU multivariate signals, respectively, where $T$ represents temporal length, $L$ denotes ECG leads, and $K$ represents ICU physiological variables. Each modality is encoded using a dedicated deep encoder $E_m(\cdot)$, producing latent embeddings $F_m \in R^d$. These embeddings are then fused using an adaptive attention mechanism $F(\cdot)$, followed by domain-invariant projection $\varphi(\cdot)$, and finally mapped to disease predictions through classifier $C(\cdot)$. The full transformation is defined as:

$$F_{ecg} = E_{ecg}(X^{ecg}), F_{ppg} = E_{ppg}(X^{ppg}), F_{icu} = E_{icu}(X^{icu}) \quad (1)$$

$$F_{fusion} = F(F_{ecg}, F_{ppg}, F_{icu}), Z = \phi(F_{fusion}), \hat{y} = C(Z) \quad (2)$$

Here, $E_m(\cdot)$ denotes modality-specific encoders, $F(\cdot)$ denotes adaptive fusion, $\varphi(\cdot)$ represents domain-invariant projection, and $C(\cdot)$ represents the classification head. The architecture is designed to ensure robust cross-domain learning across heterogeneous clinical environments while maintaining interpretability and physiological consistency (Figure 1).

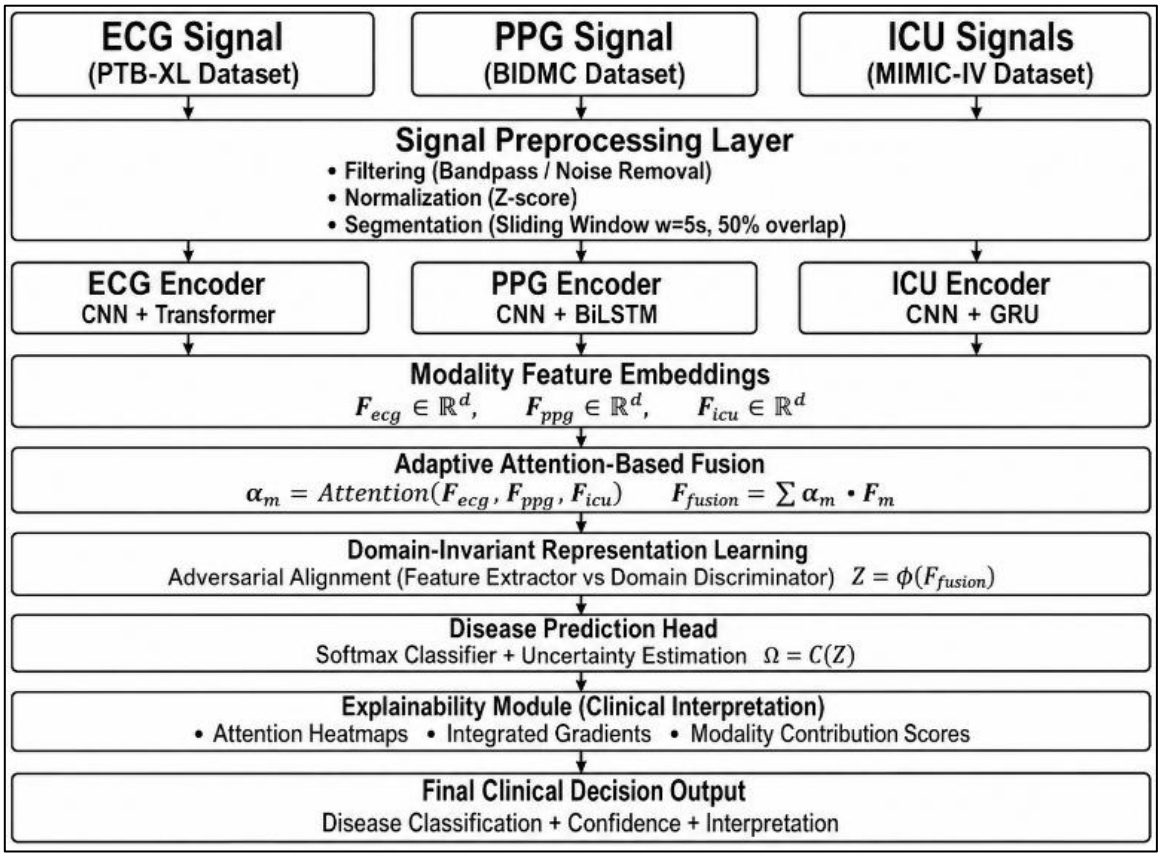


**Fig. 1.** Overview of the proposed robust cross-domain multimodal biomedical framework, illustrating modality-specific feature extraction from ECG, PPG, and ICU physiological signals, adaptive attention-based feature fusion, domain-adversarial representation learning, explainability module, and disease prediction pipeline

### *2.2 Biomedical Signal Acquisition and Preprocessing*

The study uses three PhysioNet datasets: PTB-XL (ECG: https://physionet.org/content/ptb-xl/1.0.3/), BIDMC (PPG and ICU monitoring signals: https://physionet.org/content/bidmc/1.0.0/), and MIMIC-IV (ICU clinical time-series: https://physionet.org/content/mimiciv/2.2/). All signals are standardized into a unified preprocessing pipeline to ensure cross-dataset compatibility. Table 1 summarizes the datasets used. All signals were harmonized through a unified preprocessing pipeline before multimodal training.

**Table 1** Summary of the datasets used in this study

| Dataset | Signal modality | Subjects/Records | Signal types used | Clinical labels/classes | Samples used in this study* |
|---|---|---|---|---|---|
| PTB-XL | ECG | 21,799 12-lead ECG recordings from 18,869 patients | 12-lead ECG (10 s, resampled to 250 Hz) | Diagnostic ECG categories (e.g., Normal, MI, ST/T Change, Conduction Disturbance, Hypertrophy) | After preprocessing and segmentation |
| BIDMC PPG and Respiration | PPG + bedside monitoring | 53 ICU recordings (8 min each) | PPG, ECG, impedance respiration, HR, RR, $SpO_2$ | Physiological monitoring data used for multimodal representation | After preprocessing and segmentation |
| MIMIC-IV | ICU multivariate time series | >380,000 hospital admissions from >250,000 patients (database); ICU cohort extracted for this study | Heart rate, arterial blood pressure, respiratory rate, $SpO_2$, temperature, and other routinely monitored ICU variables | Clinical outcome labels harmonized with the unified prediction task | After preprocessing and segmentation |

*The number of training samples increased after applying the sliding-window segmentation strategy.

ECG signals are resampled to 250 Hz and filtered using a 0.5–45 Hz Butterworth bandpass filter to remove baseline drift and high-frequency noise. PPG signals are filtered within 0.4–8 Hz to preserve pulse morphology. ICU time-series variables are resampled to 1-minute resolution, and missing values are handled using interpolation-based imputation.

A sliding window segmentation strategy is applied across all modalities, generating samples of fixed duration w = 5 seconds with 50% overlap. Each segment is defined as:

$$X_t = X[t:t+w] \quad (3)$$

Normalization is performed using z-score standardization:

$$X' = (X - \mu) / \sigma \quad (4)$$

where μ and σ represent the mean and standard deviation computed per subject

Disease annotations from the individual datasets were mapped to a unified diagnostic label space by grouping clinically related conditions into common prediction categories, thereby ensuring label consistency across domains. The prediction task focused on clinically relevant diagnostic categories common across the participating datasets after harmonization, enabling consistent supervised learning despite differences in the original dataset annotations. Following preprocessing, modality-specific features from all datasets were standardized using the same normalization strategy and projected into a common latent representation during model training. This harmonized preprocessing pipeline minimized dataset-specific variations while preserving modality-specific physiological characteristics required for robust cross-domain learning.

### *2.3 Modality-Specific Feature Learning*

Each modality is processed using a dedicated encoder to preserve signal-specific characteristics before multimodal fusion. The encoder architectures are designed according to the temporal and physiological characteristics of each signal while projecting all modalities into a common latent embedding space for subsequent fusion.

ECG signals are processed using a CNN–Transformer encoder comprising two 1D convolutional layers (64 and 128 filters, kernel size 5), followed by batch normalization, ReLU activation, max pooling, and a two-layer Transformer encoder (8 heads,

embedding dimension 256) to capture local morphology and long-range temporal dependencies. PPG signals are encoded using two 1D convolutional layers followed by a two-layer BiLSTM (128 hidden units) and dropout (0.3) to model pulse morphology and temporal dynamics. ICU multivariate signals are processed using a temporal CNN followed by a two-layer GRU (128 hidden units) to capture local and long-term physiological dependencies. All modality embeddings are projected into a common 256-dimensional L2-normalized latent space before attention fusion.

The latent feature representation for each modality is defined as:

$$F_m = E_m(X_m), F_m \in \mathrm{R}^d, m \in \{ecg, ppg, icu\} \quad (5)$$

where $E_m(\cdot)$ denotes the modality-specific encoder and d=256d = 256d=256 represents the common embedding dimension used for multimodal integration.

### 2.4 *Adaptive Multimodal Attention Fusion*

To handle heterogeneity and missing modalities, an adaptive attention mechanism dynamically assigns importance to each modality.

Given modality embeddings $F_m$, attention weights are computed as:

$$\alpha_m = \frac{\exp(W^T F_m)}{\sum_k \exp(W^T F_k)} \quad (6)$$

where $W$ is a learnable parameter vector. The fused representation is computed as:

$$F_{fusion} = \sum_m \alpha_m F_m \quad (7)$$

This mechanism allows the model to suppress noisy or unreliable modalities, which is critical in real-world clinical environments.

### 2.5 *Domain-Invariant Representation Learning*

To address distribution shifts across datasets, a domain-adversarial learning strategy is employed. The fused representation is mapped into a latent space:

$$Z = \varphi(F_{fusion}) \quad (8)$$

A domain discriminator D(·) predicts dataset origin d, while the feature extractor is trained adversarially:

$$\min_{\varphi, C} \max_{D} L_{task}(C(Z), y) - \lambda L_{domain}(D(Z), d) \quad (9)$$

where y denotes disease labels, d denotes domain labels, and λ controls the trade-off between task and domain invariance learning.

This formulation enforces domain-invariant yet class-discriminative representations.

### 2.6 *Disease Prediction and Explainability*

The latent embedding Z is passed through a fully connected classifier:

$$\hat{y} = \mathrm{softmax}(W_2\, \sigma(W_1\, Z + b_1) + b_2) \quad (10)$$

where $W_1$, $W_2$ and $b_1$, $b_2$ are learnable parameters, and σ(·) is ReLU activation.

Prediction confidence is estimated using the entropy of the softmax output to support uncertainty-aware decision-making.

For interpretability, Integrated Gradients are applied for feature attribution, and modality importance is computed as:

$$C_m = \frac{1}{T} \sum_{t=1}^{T} \alpha_m^t \quad (11)$$

where $C_m$ represents the contribution of modality $m$ over time.

### 2.7 *Experimental Setup*

All models were trained under identical preprocessing, dataset partitions, optimization settings, and evaluation metrics. Training used AdamW (learning rate 2×10−5, batch size 32, epochs 25, embedding dimension 256, dropout 0.3–0.5). Hyperparameters were selected through preliminary validation experiments, and reported results represent average performance across the adopted evaluation protocol.

## 3. RESULTS AND DISCUSSION

### 3.1 *Overall Disease Prediction Performance*

The proposed framework was evaluated against CNN, LSTM, Transformer, Tensor Fusion Networks, and a recent multimodal attention-based model using identical preprocessing, dataset splits, and optimization settings on the PTB-XL, BIDMC, and MIMIC-IV datasets. As shown in Table II, the proposed method consistently outperformed all baselines across all metrics, with the largest gains observed in F1-score and ROC-AUC, demonstrating improved class separability and robustness under imbalanced and clinically heterogeneous conditions.

**Table 2** Performance Comparison With State-Of-The-Art Methods

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-score (%) | ROC-AUC (%) |
|---|---|---|---|---|---|
| CNN (1D) | 84.2 | 82.9 | 81.5 | 82.1 | 86.4 |
| LSTM | 86.7 | 85.1 | 84.3 | 84.6 | 88.2 |
| Transformer | 89.4 | 88.0 | 87.6 | 87.8 | 91.5 |
| Tensor Fusion Network | 90.2 | 89.1 | 88.7 | 88.9 | 92.3 |
| Multimodal Attention Network | 91.6 | 90.4 | 90.1 | 90.2 | 93.7 |
| Proposed Framework | 94.8 | 93.9 | 93.5 | 93.7 | 96.1 |

### 3.2 *Cross-Domain Generalization Analysis*

Robustness under domain shift was evaluated through cross-domain experiments in which models trained on one dataset were tested on unseen datasets without fine-tuning. As shown in Table III, all baseline models exhibited substantial performance degradation, particularly CNN and LSTM, whereas the proposed framework consistently achieved the best cross-domain performance. The average improvement of 7–10% over Tensor Fusion Networks demonstrates the effectiveness of adversarial domain alignment in learning domain-invariant representations across different datasets.

**Table 3** Cross-domain performance (F1-score %)

| Training → Testing | CNN | LSTM | Transformer | Tensor Fusion | Proposed |
|---|---|---|---|---|---|
| PTB-XL → BIDMC | 62.3 | 65.8 | 69.4 | 71.2 | 78.6 |
| PTB-XL → MIMIC-IV | 60.7 | 64.1 | 68.2 | 70.5 | 77.3 |
| BIDMC → PTB-XL | 63.5 | 66.9 | 70.1 | 72.4 | 79.1 |
| MIMIC-IV → PTB-XL | 61.8 | 65.2 | 69.0 | 71.0 | 78.0 |

### 3.3 *Robustness Under Noisy Signal Conditions*

Robustness to noisy clinical environments was evaluated by injecting Gaussian noise at different SNR levels. As shown in Fig. 2, all models exhibited performance degradation with increasing noise, although the proposed framework consistently maintained the highest F1-score. This improved robustness is attributed to adaptive attention fusion, which down-weights corrupted modalities, and domain-invariant representation learning, which stabilizes predictions under noisy conditions.

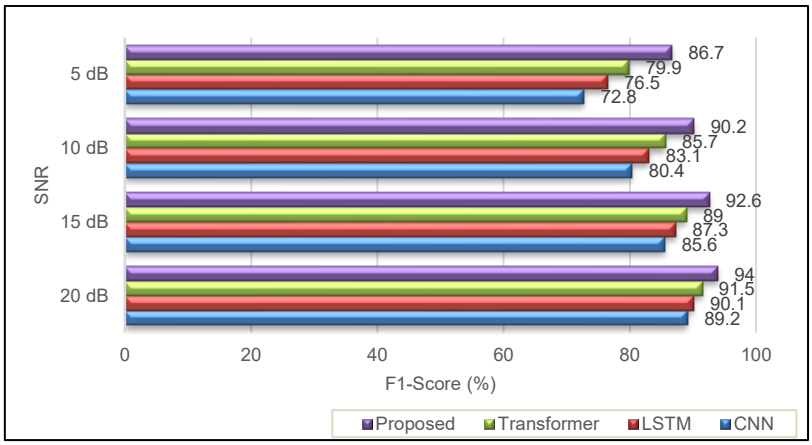


**Fig. 2.** Comparison of F1-score performance of the proposed framework and baseline models under different signal-to-noise ratio (SNR) levels, demonstrating the robustness of the proposed method against increasing noise.

### *3.4* Missing-*Modality Robustness Evaluation*

Missing modalities are common in clinical practice because of sensor failure or incomplete monitoring. As shown in Fig. 3, the proposed framework consistently outperformed baseline models under all modality availability scenarios by adaptively redistributing attention across the remaining signals. The gradual performance degradation demonstrates robust multimodal representations that are resilient to sensor dropout, making the framework well suited for real-world clinical deployment.

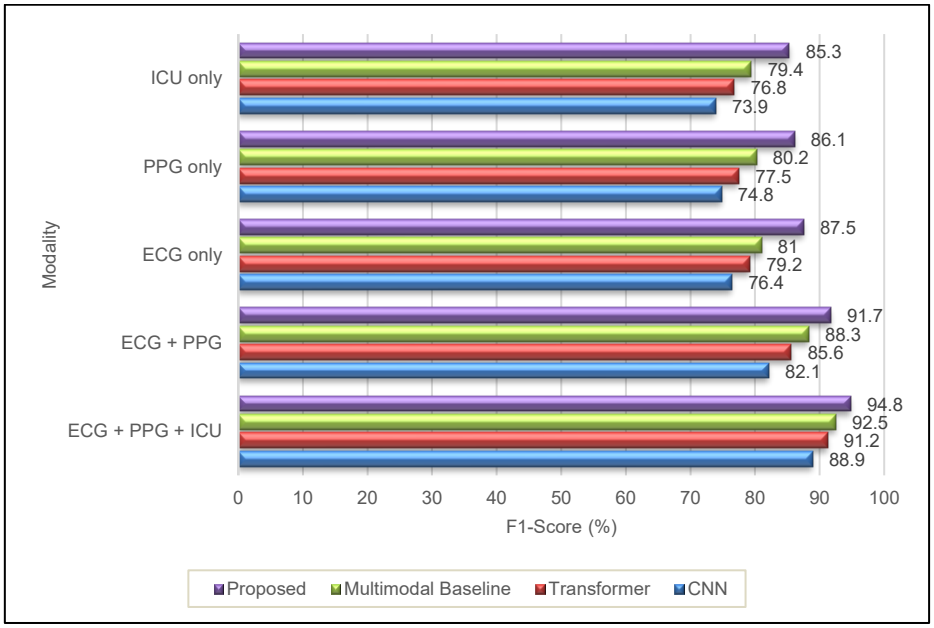


**Fig. 3.** Performance comparison under different modality availability scenarios, showing the ability of the proposed framework to maintain robust prediction accuracy despite missing ECG, PPG, or ICU signal modalities

### 3.5 Explainability and Clinical Interpretation

Explainability was achieved through attention visualization and Integrated Gradients, providing both modality- and feature-level interpretations. As shown in Figs. 4 and 5, the learned modality contributions and feature attributions align with known clinical physiology, highlighting informative regions such as ECG R-peaks, PPG systolic peaks, and respiratory phases. These results demonstrate that the proposed framework learns clinically meaningful multimodal representations, improving the transparency and trustworthiness of its predictions.

Natural Resources for Human Health
eISSN: 2583-1194

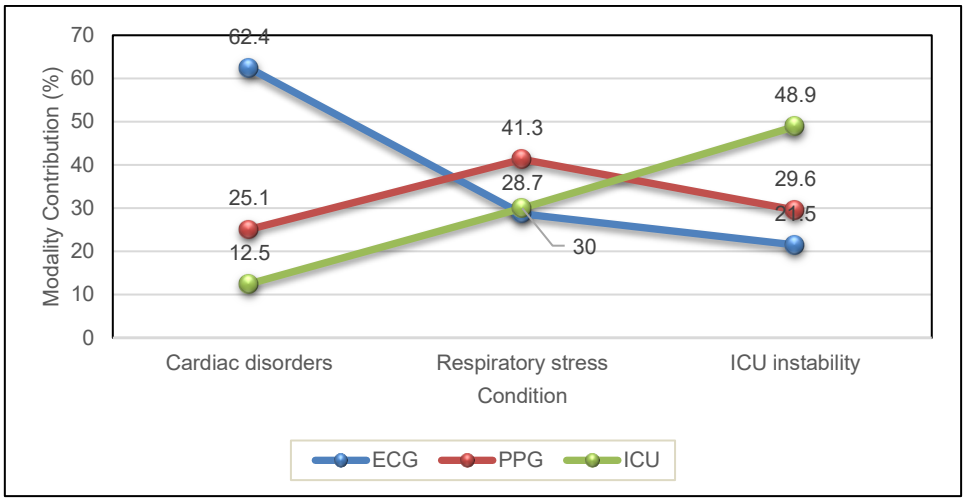


**Fig. 4.** Relative contribution of ECG, PPG, and ICU physiological signals to disease prediction across different clinical conditions, illustrating the modality-level interpretability provided by the proposed framework

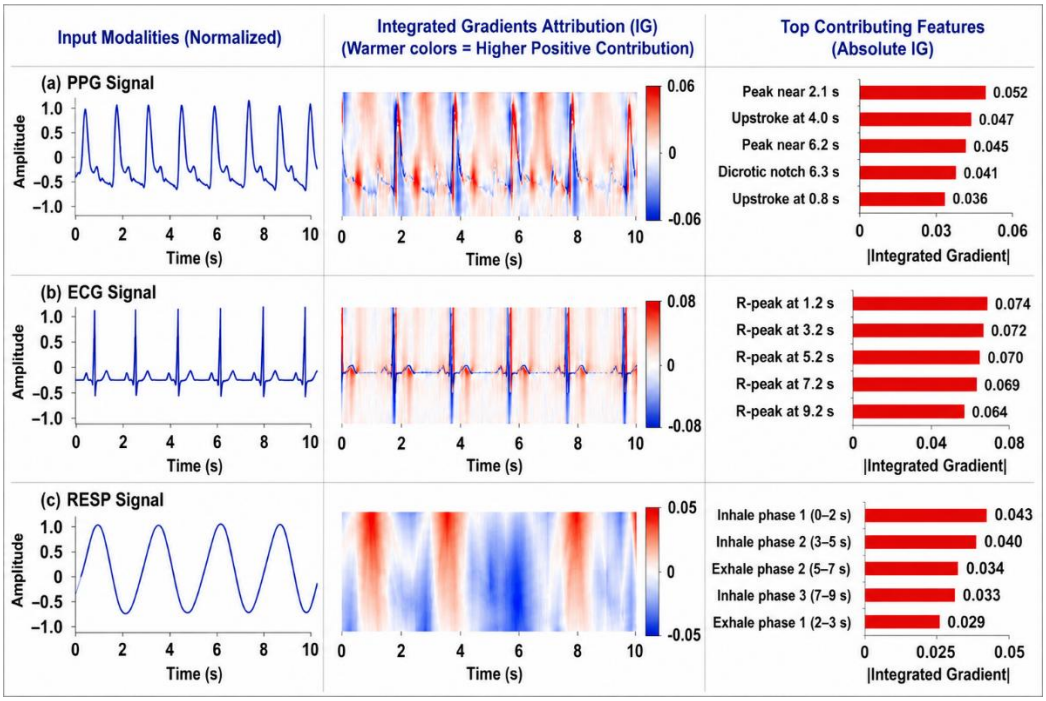


**Fig. 5.** Integrated Gradients visualization showing feature contributions for representative multimodal samples. Warmer colors indicate features with higher attribution scores.

### 3.6 Ablation Study

Table 4 shows that removing domain adaptation caused the largest performance drop (−6.9%), followed by adaptive fusion (−4.6%), confirming their critical roles in improving cross-domain robustness and multimodal learning. In contrast, removing the explainability module had a negligible impact on predictive performance, indicating that interpretability is achieved without compromising accuracy.

**Table 4** Ablation study (F1-score %)

| Configuration | F1-score (%) | Drop |
|---|---|---|
| Full Framework | 93.7 | — |
| w/o Adaptive Fusion | 89.1 | -4.6 |
| w/o Domain Adaptation | 86.8 | -6.9 |
| w/o Explainability | 93.5 | -0.2 |

### 3.7 Computational Complexity Analysis

Table 5 compares the computational efficiency of the evaluated models. The proposed framework requires lower computational cost than Transformer and Tensor Fusion models while achieving superior predictive performance, providing an effective balance between accuracy and efficiency for near real-time clinical applications.

**Table 5** Computational Efficiency

| Model | Params (M) | FLOPs (G) | Train Time (hrs) | Inference (ms) |
|---|---|---|---|---|
| CNN | 2.1 | 1.8 | 4.2 | 12.5 |
| LSTM | 3.8 | 2.9 | 6.1 | 18.3 |
| Transformer | 6.5 | 5.4 | 8.7 | 21.9 |
| Tensor Fusion | 7.2 | 6.1 | 9.4 | 24.6 |
| Proposed | 5.3 | 4.2 | 7.2 | 16.4 |

### 3.8 Statistical Significance Analysis

Statistical validation using paired t-test and Wilcoxon signed-rank test confirms that improvements are not due to random variation. Paired t-test and Wilcoxon signed-rank analyses confirmed that performance improvements were statistically significant ($p < 0.01$). The extremely low p-values confirm that performance improvements are statistically significant and consistent across folds.

### 3.9 Discussion

The proposed framework consistently outperformed conventional CNN, LSTM, Transformer, and Tensor Fusion models, demonstrating superior predictive accuracy, robustness, and cross-domain generalization. Compared with recent studies using similar PhysioNet datasets, the proposed model achieved 94.8% accuracy, 93.7% F1-score, and 96.1% ROC-AUC, exceeding the 91.5% ROC-AUC reported by Xia et al. [5] while extending beyond single-modality ECG analysis. Similarly, unlike the ECG–PPG Conv-LSTM framework of Kamanditya et al. [7], the proposed approach integrates ECG, PPG, and ICU physiological signals with domain-adversarial learning, improving robustness to unseen domains. It also outperformed the multimodal MIMIC-IV framework of Wang et al. [18] (AUROC 0.916) while maintaining robustness to missing modalities. Although ECG-XPLAIM [19] achieved higher AUROC for specific arrhythmia detection, it is limited to ECG analysis, whereas the proposed framework generalizes across heterogeneous multimodal datasets. Furthermore, attention visualization and Integrated Gradients provide clinically meaningful explanations, enhancing the transparency of AI-assisted decision-making. Despite these promising results, further validation on larger multi-center datasets, together with unsupervised domain adaptation and lightweight deployment, would further improve clinical applicability.

## 4. CONCLUSIONS

Reliable interpretation of heterogeneous biomedical signals remains challenging because of dataset variability, noisy inputs, and missing modalities. To address these issues, this study proposed a robust cross-domain multimodal deep learning framework integrating modality-specific feature learning, adaptive attention-based fusion, domain-adversarial learning, and explainability for disease prediction using ECG, PPG, and ICU physiological signals. Evaluation on the PTB-XL, BIDMC, and MIMIC-IV datasets achieved 94.8% accuracy, 93.7% F1-score, and 96.1% ROC-AUC, consistently outperforming CNN, LSTM, Transformer, and tensor fusion baselines while improving cross-domain robustness. These results demonstrate that adaptive multimodal fusion with domain-invariant learning provides a reliable and clinically interpretable solution for biomedical signal analysis. Future work will focus on unsupervised domain adaptation, lightweight optimization, and validation on larger multi-center clinical datasets.

## AUTHOR CONTRIBUTIONS

Conceptualization, Dr. Suvarna Sunil Nirmal and Dr. Sneha Arjun Khaire; methodology, Dr. Suvarna Sunil Nirmal; software, Rageshri Bakare; validation, Anuradha S. Deshpande and Dr. Minal Vilas Gade; formal analysis, Dr. Suvarna Sunil Nirmal; investigation, Dr. Sneha Arjun Khaire; resources, Amit Kurmi; data curation, Rageshri Bakare; writing original draft preparation, Dr. Suvarna Sunil Nirmal; writing review and editing, Dr. Sneha Arjun Khaire and Anuradha S. Deshpande; visualization, Dr. Minal Vilas Gade; supervision, Dr. Sneha Arjun Khaire; project administration, Amit Kurmi; funding acquisition, Dr. Minal Vilas Gade. All authors have read and agreed to the published version of the manuscrip.

## FUNDING

This research received no external funding.

## DATA AVAILABILITY STATEMENT

The datasets used in this study are publicly available from the respective sources cited in the manuscript. The data supporting the findings of this study can be obtained from these publicly accessible datasets. No new datasets were generated during the study.

## CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests.

## ACKNOWLEDGEMENT

This research was not funded by any grant.

## REFERENCES

[1] N. Naik, N. Kassetty, S. Chippagiri, P. Randhawa, B. M. Zeeshan Hameed, and V. Patil, "Design and optimisation of an IoT-based artificial intelligence framework for real-time health monitoring and telemedicine diagnostics in smart healthcare systems," Discov. Internet Things, vol. 6, no. 1, p. 37, Feb. 2026, doi: 10.1007/S43926-026-00306-W.

[2] Y. J. Lee, C. Park, H. Kim, S. J. Cho, and W. H. Yeo, "Artificial intelligence on biomedical signals: technologies, applications, and future directions," Med-X, vol. 2, no. 1, p. 25, Dec. 2024, doi: 10.1007/S44258-024-00043-1.

[3] A. Kumar, E. Tripathi, A. K. Tripathi, H. K. Diwedi, P. S. Rathore, and A. S. Ansari, "SBTM: epileptic seizure prediction from EEG signal using deep learning in blockchain-enabled smart healthcare monitoring with IoT networking," Sci. Rep., vol. 16, no. 1, p. 6830, Jan. 2026, doi: 10.1038/s41598-026-36425-1.

[4] S. Gore, A. Salati, P. Satish Kumar, J. Keerthika, M. Y. Pasha, and J. Yogapriya, "Title of the paper," in Proc. 6th Int. Conf. Electron. Sustain. Commun. Syst. (ICESC), Coimbatore, India, Sep. 2025, pp. 1428–1435, doi: 10.1109/ICESC65114.2025.11212399.

[5] Y. Xia, Y. Xiong, and K. Wang, "A transformer model blended with CNN and denoising autoencoder for inter-patient ECG arrhythmia classification," Biomed. Signal Process. Control, vol. 86, p. 105271, Sep. 2023, doi: 10.1016/J.BSPC.2023.105271.

[6] A. Kline et al., "Multimodal machine learning in precision health: A scoping review," npj Digit. Med., vol. 5, no. 1, pp. 171-, Nov. 2022, doi: 10.1038/s41746-022-00712-8.

[7] B. Kamanditya, Y. N. Fuadah, N. Q. T. Mahardika, and K. M. Lim, "Continuous blood pressure prediction system using Conv-LSTM network on hybrid latent features of photoplethysmogram (PPG) and electrocardiogram (ECG) signals," Sci. Rep., vol. 14, no. 1, pp. 16450-, Jul. 2024, doi: 10.1038/s41598-024-66514-y.

[8] L. Monish and S. G. Shaila, "A Hybrid Fuzzy CNN–LSTM Approach for Emotion Recognition from EEG–ECG Physiological Signals", Eng. Technol. Appl. Sci. Res., vol. 15, no. 6, pp. 30405–30411, Dec. 2025. https://doi.org/10.48084/etasr.14864.

[9] S. Gore, S. Jadhav, S. Gore, K. D. Deshmukh, N. Rane, and P. Rao, "Title of the paper," in Lecture Notes in Networks and Systems, vol. 1600, A. Swaroop, B. Virdee, S. D. Correia, and Z. Polkowski, Eds. Cham, Switzerland: Springer, 2026, pp. 312–325, doi: 10.1007/978-3-032-03072-6_25.

[10] S. Liang, L. Li, W. Zu, W. Feng, and W. Hang, "Adaptive deep feature representation learning for cross-subject EEG decoding," BMC Bioinformatics, vol. 25, no. 1, pp. 393-, Dec. 2024, doi: 10.1186/S12859-024-06024-W.

[11] I. Jemal, L. Abou-Abbas, K. Henni, A. Mitiche, and N. Mezghani, "Domain adaptation for EEG-based, cross-subject epileptic seizure prediction," Front. Neuroinform., vol. 18, p. 1303380, Feb. 2024, doi: 10.3389/FNINF.2024.1303380.

[12] S. Liu, Z. Wang, Y. An, B. Li, X. Wang, and Y. Zhang, "DA-CapsNet: A multi-branch capsule network based on adversarial domain adaption for cross-subject EEG emotion recognition," Knowledge-Based Syst., vol. 283, p. 111137, Jan. 2024, doi: 10.1016/J.KNOSYS.2023.111137.

[13] H. Cen, M. Zhao, K. Cui, F. Tian, Q. Zhao, and B. Hu, "A convolution and attention-based conditional adversarial domain adaptation neural network for emotion recognition using electroencephalography," Biomed. Signal Process. Control, vol. 100, p. 106957, Feb. 2025, doi: 10.1016/J.BSPC.2024.106957.

[14] A. G. Khoee, Y. Yu, and R. Feldt, "Domain generalization through meta-learning: a survey," Artif. Intell. Rev., vol. 57, no. 10, pp. 285-, Sep. 2024, doi: 10.1007/S10462-024-10922-Z.

[15] K. Chen, E. Gal, H. Yan, and H. Li, "Domain Generalization with Small Data," Int. J. Comput. Vis., vol. 132, no. 8, pp. 3172–3190, Mar. 2024, doi: 10.1007/S11263-024-02028-4.

[16] Z. Lou, R. Wang, B. E. Y. Belmekki, and M.-S. Alouini, "Towards bioelectric signal-enabled human healthcare monitoring: state-of-the-art, design strategies, challenge, and future," npj Biomed. Innov., vol. 3, no. 1, pp. 7-, Jan. 2026, doi: 10.1038/s44385-025-00061-7.

[17] M. G. Garcés, J. C. Montoya, M. I. P. Martínez, J. V. García, and E. H. H. Rincón, "Deep Learning and Noninvasive Sensors for Detecting Physiological Dysregulation: A Scoping Review," J. Med. Syst., vol. 50, no. 1, pp. 14-, Jan. 2026, doi: 10.1007/S10916-025-02332-7.

[18] M. Wang, S. Fan, Y. Li, Z. Xie, and H. Chen, "Missing-modality enabled multi-modal fusion architecture for medical data," Journal of Biomedical Informatics, vol. 164, Art. no. 104796, Apr. 2025, doi: 10.1016/j.jbi.2025.104796.

[19] P. Pantelidis, S. Ruipérez-Campillo, J. E. Vogt, A. Antonopoulos, I. Gialamas, G. E. Zakynthinos, M. Spartalis, P. Dilaveris, J. Millet, T. G. Papaioannou, E. Oikonomou, and G. Siasos, "ECG-XPLAIM: EXPlainable locally-adaptive artificial intelligence model for arrhythmia detection from large-scale electrocardiogram data," Frontiers in Cardiovascular Medicine, vol. 12, Art. no. 1659971, 2025, doi: 10.3389/fcvm.2025.1659971.