A Survey of ECG Multimodal Representation Learning for Cardiovascular Health Enhancement

Abstract

This survey explores the integration of multimodal representation learning into electrocardiogram (ECG) analysis, highlighting its transformative potential for enhancing cardiovascular diagnostics. By synthesizing diverse data modalities, such as ECG signals, phonocardiograms, cardiac magnetic resonance images, and clinical text reports, this approach provides a comprehensive framework for understanding cardiac function. Advanced machine learning techniques, including deep learning, self-supervised learning, and transfer learning, address the limitations of traditional ECG analysis methods, which often rely on manual interpretation and extensive labeled data. The integration of multimodal data sources with ECG signals significantly enhances diagnostic accuracy and interpretability, enabling healthcare practitioners to achieve a holistic understanding of cardiovascular health, leading to improved diagnostic outcomes and personalized treatment strategies. The survey concludes by emphasizing the importance of continued research and development to further enhance the accuracy, efficiency, and robustness of ECG analysis, particularly in resource-limited settings. Future advancements hold significant promise for revolutionizing cardiovascular health monitoring and treatment, ultimately leading to better patient outcomes and more effective healthcare delivery.

1 Introduction

1.1 Significance of ECG in Cardiovascular Health

The electrocardiogram (ECG) is vital for diagnosing and monitoring cardiovascular diseases (CVDs), responsible for over 17 million deaths annually [1]. As a non-invasive diagnostic tool, the ECG provides essential insights into cardiac structure and electrical activity, making it the most commonly performed test in clinical practice [2]. Its significance in early detection and management of cardiovascular conditions is critical, enabling healthcare professionals to identify various cardiac abnormalities, including arrhythmias and myocardial infarctions [3].

Accurate ECG interpretation is crucial for effective cardiac disease management, yet challenges persist. Even seasoned cardiologists may struggle with complex arrhythmias like atrial fibrillation (AF), which increases stroke and mortality risk. ECGs are also essential in hospital settings for continuous monitoring, where abnormal heart rhythms can indicate severe conditions, including sudden cardiac death [4].

In addition to traditional applications, ECG data are increasingly used to diagnose non-cardiac conditions, such as liver diseases, by exploiting the physiological links between cardiovascular and hepatic health [5]. In resource-limited settings, ECGs provide a practical and cost-effective solution for ongoing patient assessment, significantly aiding in CVD prevention [6]. The ECG's enduring relevance in cardiovascular health management underscores its role as a cornerstone for early detection and effective management of cardiac conditions, ultimately improving patient outcomes [7].

Despite its widespread application, the classification of CVDs using ECGs, especially when stored as images, faces limitations that require attention [8]. A thorough understanding of ECG monitoring

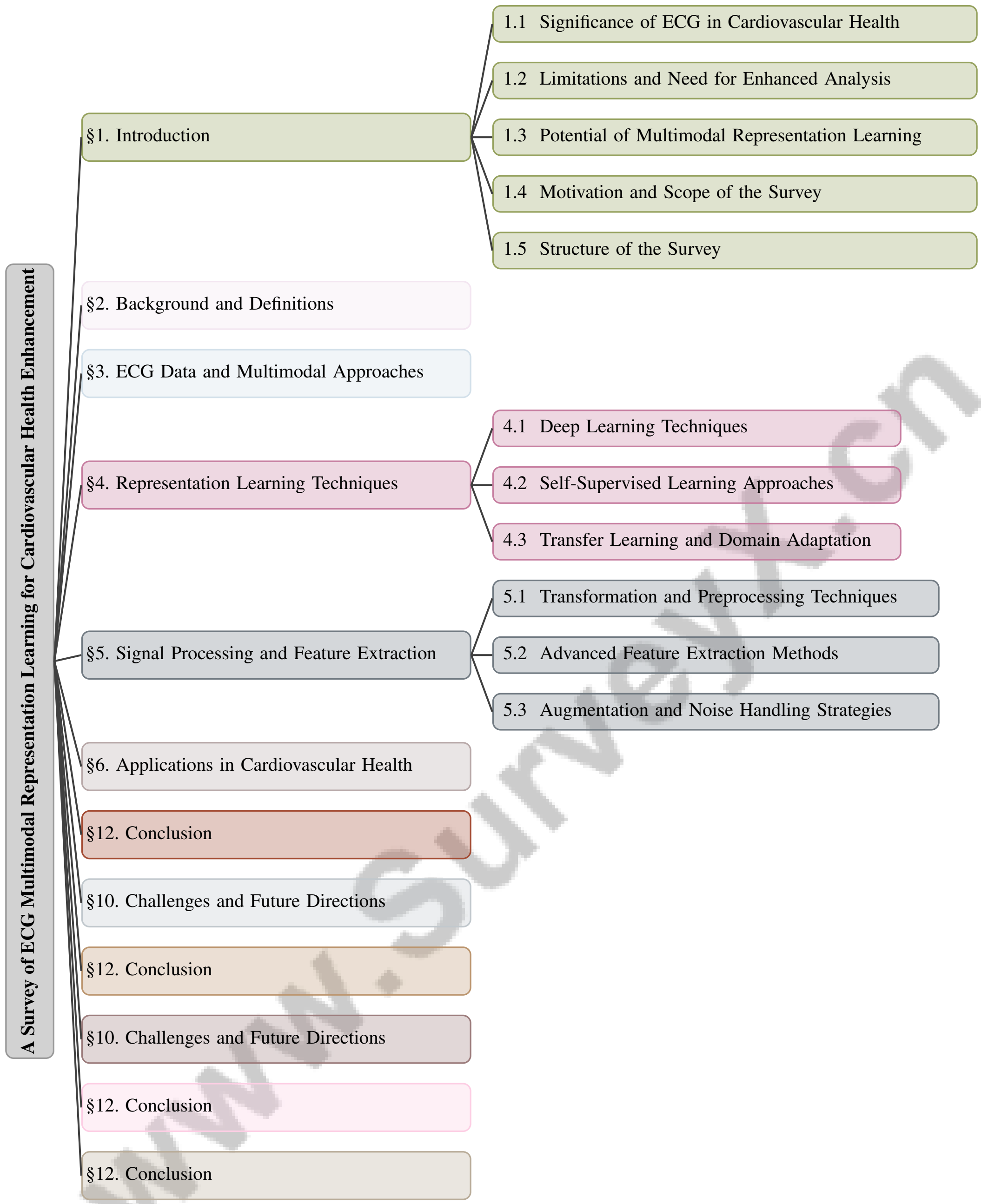


Figure 1: chapter structure

systems is essential to address these challenges, as highlighted in recent surveys focusing on the selection, comparison, and evaluation of various systems [9].

1.2 Limitations and Need for Enhanced Analysis

Current ECG analysis methods face significant limitations that undermine their diagnostic efficacy and clinical applicability, highlighting the need for advanced analytical techniques. A primary challenge is the reliance on manual interpretation, which is often inconsistent and time-consuming, leading to potential human errors in detecting complex arrhythmias [10]. This issue is exacerbated by imbalanced ECG datasets, variability in signal lengths, and extensive training data requirements, which hinder performance in real-time applications [11].

AI-generated, for reference only.

Furthermore, existing ECG analysis techniques typically depend on all 12 leads, posing a significant limitation for healthcare facilities lacking complete lead sets [12]. This dependence complicates diagnostics in resource-constrained settings, emphasizing the need for methodologies that effectively utilize fewer leads without sacrificing accuracy. Additionally, the scarcity of labeled data and the co-occurrence of multiple cardiovascular diseases present challenges, leading to poor performance on unseen datasets [13].

The advancement of automatic ECG interpretation has been hindered by insufficient datasets and poorly defined evaluation procedures, which are crucial for comparing and validating different algorithms [14]. The rapid evolution of technologies and the vast array of existing systems lack comprehensive frameworks for understanding and evaluating these systems, complicating the selection and implementation of effective ECG analysis solutions [9].

Machine learning and deep learning techniques show promise in addressing some of these challenges by enabling automatic feature extraction from ECG data [15]. However, these methods encounter obstacles, including the poor performance of existing discriminative models under limited channel conditions and the inability of techniques like -VAEs to adequately capture task-specific information. Additionally, the computational overhead associated with traditional encryption methods compromises real-time processing capabilities essential for immediate medical responses [16].

These limitations underscore the urgent need for innovative analytical techniques that enhance the accuracy, efficiency, and interpretability of ECG analysis, ultimately leading to improved cardiovascular health outcomes across diverse clinical settings [17].

1.3 Potential of Multimodal Representation Learning

Multimodal representation learning offers significant potential in overcoming the limitations of traditional ECG analysis by integrating diverse data modalities and employing advanced learning frameworks. This approach enhances both diagnostic accuracy and interpretability. For example, combining knowledge-based abductive interpretation with machine learning classifiers has been shown to improve ECG classification accuracy [18]. Additionally, channel-wise attention mechanisms within convolutional neural networks effectively leverage information from multiple ECG leads, addressing constraints faced by conventional methods [19].

The hybrid application of 1D and 2D convolutional neural networks (CNNs) has yielded significant improvements in ECG signal classification accuracy by harnessing deep learning strengths [1]. Generative modeling approaches have also been adapted to enhance ECG classification performance, even with limited channel availability, showcasing the versatility of generative models across domains [20]. These methodologies highlight multimodal learning frameworks' potential to address traditional ECG analysis limitations.

Innovations in self-supervised learning have been pivotal in enhancing model understanding of dynamic ECG signal characteristics without extensive data augmentation, thus increasing the robustness of ECG analysis [21]. The development of artificial ECG models simulating PTSD-related abnormalities has further improved classifier performance and broadened the understanding of ECG data in various health contexts [22].

The promise of multimodal representation learning is exemplified by contrastive learning approaches that align representations of incomplete ECG signals with those of complete signals, thereby improving diagnostic performance with fewer leads [12]. Furthermore, chaotic encryption techniques present a viable alternative for enhancing ECG analysis by offering high security and efficiency due to their unpredictability and sensitivity to initial conditions [16].

Moreover, semi-supervised learning (SSL) methods, such as ECGMatch, utilize unlabeled data to enhance model performance, addressing challenges posed by the scarcity of labeled data [13]. Utilizing large datasets with predefined train-test splits, as introduced in benchmarking studies, improves the comparability and reliability of results, facilitating robust ECG analysis model development [14].

By leveraging these advanced frameworks, multimodal representation learning significantly enhances the precision, interpretability, and comprehensiveness of cardiovascular diagnostics, leading to improved patient outcomes and more effective ECG analysis. This underscores the transformative potential of multimodal approaches in cardiovascular health monitoring, as evidenced by innovative techniques that integrate ECG data, clinical reports, and patient metadata. Utilizing advanced machine learning methodologies, such as vision-language learning and contrastive learning, these approaches not only improve diagnostic accuracy but also facilitate the retrieval of similar clinical cases, benefiting resource-limited settings. Moreover, integrating diverse data sources, including cardiac magnetic resonance images and ECG signals, fosters a more comprehensive understanding of patients' cardiovascular health, ultimately leading to improved prediction models and early intervention strategies [23, 12, 24].

1.4 Motivation and Scope of the Survey

This survey is motivated by the pressing need to enhance ECG diagnostic accuracy by integrating multimodal data sources, such as clinical text, with ECG signals. This strategy aims to overcome current methodologies' limitations in accurately diagnosing a wide range of cardiac abnormalities, particularly in short-duration 12-lead ECGs. The survey will explore advanced techniques, including large-scale cross-modality pretrained models like CardiacNets, which could significantly improve ECG analysis and diagnostic accuracy [25].

Another critical motivation is to investigate age-related changes in ECG data among healthy individuals, distinguishing normal aging effects from those caused by cardiovascular diseases [26]. This differentiation is essential for developing precise diagnostic tools capable of effectively identifying pathological changes in ECG signals.

The survey also addresses the need for effective ambulatory monitoring systems and advanced ECG signal processing techniques, which are crucial for accurate and efficient ECG classification in various clinical settings [27]. Outdated ECG interpretation methods that do not leverage modern technological advancements limit diagnostic capabilities and lack interpretability in automated systems [2]. By exploring innovative approaches, this survey aims to overcome these limitations and improve ECG analysis.

The scope of this survey encompasses a comprehensive examination of recent advancements in computer-aided ECG analysis, focusing on their applicability in real-world medical diagnostics. It will cover a wide range of ECG monitoring systems, including their design, classification, and analysis, and seek to fill knowledge gaps by proposing a verified taxonomy and architectural model for these systems [9]. The survey aims to enhance the accuracy of automatic ECG analysis, particularly in settings with limited access to specialized cardiologists, by addressing challenges posed by biased cardiac abnormality distributions in datasets [3].

1.5 Structure of the Survey

This survey is systematically structured to provide a comprehensive exploration of ECG multimodal representation learning and its applications in enhancing cardiovascular health. The paper begins with an **Introduction**, establishing the significance of ECG in cardiovascular health monitoring, discussing current limitations in ECG analysis methods, and highlighting the potential of multimodal representation learning to address these challenges. The introduction concludes by outlining the survey's motivation and defining its scope, laying the groundwork for an in-depth exploration of various aspects of ECG analysis, including preprocessing, feature extraction, and classification techniques, and their implications for diagnosing cardiac conditions and enhancing clinical decision-making [28, 29, 30, 31, 32].

The second section, **Background and Definitions**, provides foundational knowledge of key concepts such as ECG, multimodal data integration, representation learning techniques, signal processing, cardiovascular health, and feature extraction, ensuring readers are equipped with necessary terminology and context for subsequent discussions.

In the **ECG Data and Multimodal Approaches** section, the nature of ECG data is explored, emphasizing its role in cardiovascular health monitoring. The integration of multimodal data sources with ECG is discussed, highlighting the benefits and challenges of using multiple data types for comprehensive health analysis.

The fourth section, **Representation Learning Techniques**, delves into various techniques applied to ECG data, including deep learning, self-supervised learning, and transfer learning, showcasing their applications in enhancing feature extraction and signal processing for ECG analysis.

The fifth section focuses on **Signal Processing and Feature Extraction**, detailing the techniques used in ECG analysis and discussing crucial feature extraction methods for identifying cardiovascular health indicators. This section also explores how advanced algorithms improve the accuracy and reliability of ECG interpretations.

The practical applications of ECG multimodal representation learning in cardiovascular health are presented in the **Applications in Cardiovascular Health** section, including case studies and examples where these techniques have enhanced diagnosis, monitoring, or treatment of cardiovascular conditions.

In the seventh section, **Challenges and Future Directions**, current challenges in implementing multimodal representation learning for ECG data are identified, along with potential future directions and research opportunities to overcome these challenges and improve cardiovascular health outcomes.

Finally, in the **Conclusion**, we synthesize the primary findings of the survey, emphasizing the critical role of integrating multimodal representation learning in ECG analysis for enhancing cardiovascular health diagnostics. This integration addresses the challenge of limited access to complete ECG lead sets by improving diagnostic accuracy through techniques like contrastive learning, while also highlighting the transformative potential of advanced methodologies such as Large Language Models and Vision-Transformer models. Such advancements could significantly enhance automated ECG interpretation and retrieval systems, ultimately improving clinical decision-making and accessibility, especially in underdeveloped regions [12, 24].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Electrocardiogram (ECG) and Its Role in Cardiovascular Health

The electrocardiogram (ECG) is a crucial non-invasive diagnostic tool in cardiovascular health, providing insights into the heart's electrical activity and aiding in the detection of conditions like arrhythmias and myocardial infarctions. Utilizing a 12-lead system, ECGs enable thorough evaluations of cardiac function, including heart rate and rhythm analysis, which are vital for diagnosing various heart conditions. ECG applications extend beyond traditional uses to areas such as emotion recognition and biometric identification [33, 28, 34, 35, 32].

The ability of ECGs to detect complex wave segments and their temporal relationships is essential for diagnosing intricate heart conditions. Extensive ECG datasets, particularly those annotated with the SCP-ECG standard, enhance research capabilities in predicting ECG outcomes and improve diagnostic accuracy through advanced automated analysis methods. These methods include deep learning models that generate clinician-like interpretations and connect ECG signals with textual reports. Resources like the ECG-Image-Database support the training of robust machine learning algorithms for ECG digitization and classification, advancing the diagnostic potential of ECG technology [34, 35, 36].

Despite ECG's pivotal role in diagnosing cardiovascular diseases, interpreting signals, especially for complex arrhythmias, poses challenges due to intricate waveforms and temporal variations [37]. The diagnostic accuracy of ECGs depends heavily on comprehensive lead data, as the standard 12-lead ECG provides a multidimensional view of the heart's electrical activity, crucial for effective cardiovascular health monitoring [12].

Beyond cardiac conditions, ECGs have been used to diagnose non-cardiac diseases, such as liver disorders, highlighting the physiological link between cardiac and hepatic health [5]. This underscores ECG's potential as a comprehensive diagnostic tool for both cardiac and non-cardiac conditions.

Advancements in technologies, such as representation learning and multimodal data integration, have further enhanced ECG diagnostic capabilities. By incorporating information from diverse sources, including electrocardiography signals and cardiac magnetic resonance images, these techniques provide a holistic view of cardiovascular health, facilitating the detection of complex wave segments and their temporal relationships critical for diagnosing heart conditions [37].

The development of extensive, annotated ECG datasets, particularly those adhering to the SCP-ECG standard, has been instrumental in enhancing machine learning models for ECG analysis. These datasets support the creation of advanced techniques like ECG-Text Pre-training (ETP), aligning

ECG signals with textual reports for improved representation learning, and enabling automated medical report generation that bridges signal processing and clinical interpretation. Consequently, these advancements facilitate robust zero-shot classification and automated clinical decision support, significantly progressing cardiovascular healthcare [35, 34]. The electrocardiogram remains an indispensable tool in cardiovascular health monitoring, offering critical insights into cardiac function and aiding in the early detection and management of cardiovascular diseases.

2.2 Multimodal Data Integration

The integration of multimodal data in cardiovascular diagnostics represents a significant advancement toward comprehensive health analysis. By combining diverse data modalities, such as ECG signals, phonocardiogram (PCG) signals, and clinical text reports, a nuanced understanding of cardiac health is achieved. This multifaceted approach enriches feature representation, enhancing model generalization across various diagnostic tasks and improving predictive accuracy [38].

A primary challenge in multimodal data integration is effectively translating information across different modalities, crucial for maximizing predictive capabilities [23]. This integration addresses limitations associated with single-modality analysis, such as imbalanced ECG datasets and variability in signal morphology among patients [39]. Leveraging synchronized data from multiple sources, including ECG and blood pressure (BP) signals, enables robust heartbeat detection and comprehensive health monitoring [40].

Multimodal data integration is particularly advantageous in real-time monitoring scenarios, ensuring a holistic analysis by incorporating multiple data types, thus enhancing diagnostic precision and efficiency [27]. The use of large, real-world clinical datasets, comprising over 2.3 million ECG recordings, exemplifies the potential of multimodal data integration in advancing health analysis [41]. These datasets enable the development of models that accurately reflect localized heart disease conditions in specific waveforms [42].

Moreover, CNN-based architectures with cross-attention mechanisms have been employed to fuse features from both time and frequency domains, enhancing the efficacy of multimodal integration [43]. Organizing ECG monitoring systems into various contexts, such as home, hospital, ambulatory, and remote settings, alongside technologies like IoT and mobile computing, further emphasizes the transformative potential of multimodal data integration in cardiovascular diagnostics [9].

The integration of multimodal data is pivotal in advancing cardiovascular diagnostics, offering a comprehensive framework that addresses the inherent challenges of traditional ECG analysis and supports more accurate and efficient health assessments [44].

2.3 Representation Learning Techniques

Representation learning techniques have emerged as crucial methodologies in analyzing and classifying electrocardiogram (ECG) data, significantly enhancing the accuracy and efficiency of cardiovascular diagnostics. These methods focus on extracting meaningful features from raw ECG signals, improving the interpretability and predictive capabilities of analytical models. The quasi-periodic nature of ECG signals, often accompanied by noise and artifacts, necessitates robust representation learning approaches to ensure precise signal classification [11].

Deep learning models, particularly convolutional neural networks (CNNs), have been instrumental in processing and classifying 12-lead ECG signals, yielding improved diagnostic outcomes. These models utilize hierarchical feature extraction, critical for capturing complex patterns within ECG data [9]. The integration of self-supervised learning frameworks, such as the Subject-based Non-contrastive Learning (SBnCL) approach, allows for the learning of ECG signal representations without extensive labeled data, enabling models to generalize across diverse ECG datasets with varying distributions.

Hybrid models that combine one-dimensional (1D) and two-dimensional (2D) CNNs have significantly improved ECG classification accuracy, achieving remarkable results of 97.38

The significance of representation learning extends to developing novel architectural models that categorize ECG monitoring systems into distinct layers and contexts, aiding in understanding the components and interactions within these systems [9]. By integrating multimodal data sources, such

AI-generated, for reference only.

as clinical text and other physiological signals, representation learning further enhances the potential for comprehensive cardiovascular health monitoring.

Representation learning techniques are fundamentally transforming ECG analysis by providing advanced frameworks for feature extraction that significantly improve the accuracy and interpretability of cardiovascular diagnostics. These techniques, including novel architectures such as transformers and innovative layers like the piece-wise matching layer, capture complex temporal relationships in ECG signals and facilitate the integration of cross-modal data, such as ECG signals and textual reports. Consequently, these advancements not only enhance diagnostic precision but also support the development of interpretable machine learning models that clarify decision-making processes behind ECG classifications, ultimately leading to more effective and accessible cardiovascular care [11, 45, 34, 46, 24]. These advancements hold significant promise for improving patient outcomes and advancing the capabilities of ECG-based health assessments.

3 ECG Data and Multimodal Approaches

3.1 The Nature of ECG Data

Electrocardiogram (ECG) data are essential for cardiovascular diagnostics, offering a non-invasive method to assess cardiac function by recording the heart's electrical activity through a 12-lead system, which captures signals from multiple perspectives [37]. However, ECG analysis faces challenges due to noise, artifacts, and variability in signal lengths, which can impair signal interpretation accuracy and reduce performance in real-time applications, where rapid and precise arrhythmia detection is critical [11, 10]. Additionally, the reliance on comprehensive 12-lead ECGs poses challenges in resource-limited settings lacking advanced diagnostic tools [12].

To overcome these challenges, multimodal data integration, which combines ECG data with other modalities like imaging and clinical reports, is gaining traction. This approach enhances diagnostic accuracy and supports effective clinical decision-making by refining ECG signal interpretation and aiding in retrieving relevant clinical cases [12, 24, 35, 47]. Multimodal integration allows for cross-verification of information, bolstering diagnostic model robustness and improving prediction accuracy. However, effectively combining diverse data sources remains challenging, necessitating large annotated ECG datasets, such as those adhering to the SCP-ECG standard, to advance machine learning models and enhance ECG data's diagnostic capabilities [38]. Successful multimodal data integration could revolutionize cardiovascular diagnostics by providing a more comprehensive and precise heart health assessment.

3.2 Integration of Multimodal Data Sources

Integrating multimodal data sources with ECG signals is a transformative strategy for enhancing cardiovascular diagnostics. This approach leverages diverse data types, such as cardiac magnetic resonance (CMR) imaging and clinical text, to deliver a comprehensive assessment of cardiac function and disease detection [25]. By merging ECG signals with CMR data through techniques like contrastive learning and generative pretraining, systems such as CardiacNets have improved disease detection accuracy and functional assessment, establishing a robust framework for cardiac evaluation [25].

A significant advantage of multimodal integration is its ability to facilitate high-throughput screening for various diseases based on initial in-hospital ECG readings, streamlining diagnostic processes and enhancing early disease detection [48]. Incorporating language models into ECG analysis further illustrates the promise of multimodal frameworks by enhancing the generalization performance of ECG-language question answering tasks, integrating textual data with ECG signals to provide richer diagnostic insights [49].

Advanced feature extraction techniques are crucial for effective multimodal data integration. Neural network-based methods optimize meaningful feature extraction from ECG signals, enhancing the predictive accuracy and reliability of diagnostic models [50]. These techniques synthesize information from various sources, fostering a comprehensive approach to cardiovascular health analysis.

The integration of multimodal data sources, including ECG signals and clinical reports, represents a significant advancement in cardiovascular diagnostics. It enables automated interpretation and

AI-generated, for reference only.

retrieval systems that leverage advanced machine learning techniques, such as Large Language Models and Vision-Transformer models, to improve diagnostic accuracy and accessibility, particularly in underdeveloped regions [24, 35]. By amalgamating diverse data types, healthcare practitioners can achieve more thorough and accurate assessments, ultimately leading to enhanced patient outcomes and more effective disease management strategies.

In recent years, the field of electrocardiogram (ECG) analysis has witnessed significant advancements, particularly through the application of various representation learning techniques. Figure 2 illustrates the hierarchical structure of these techniques, encompassing deep learning methods, self-supervised learning approaches, and transfer learning with domain adaptation. Each category is meticulously divided into specific methods, architectures, and applications, thereby highlighting the substantial progress made in ECG feature extraction, classification accuracy, and diagnostic performance. This structured representation not only enhances our understanding of the interrelationships among these techniques but also underscores their collective impact on improving ECG analysis outcomes.

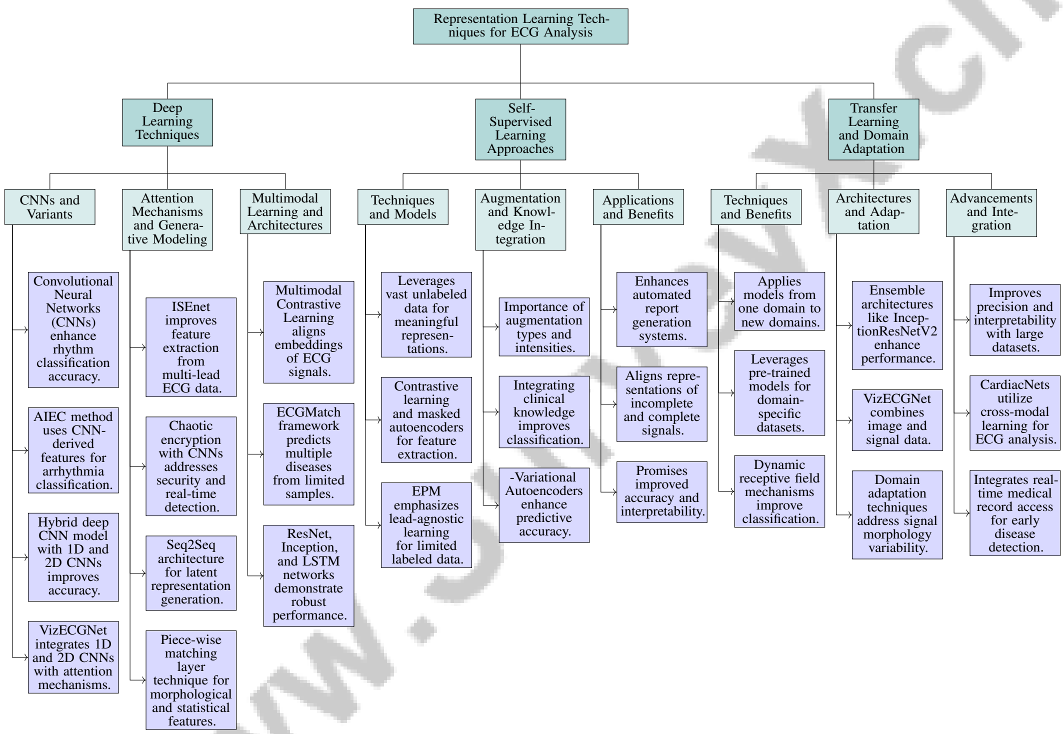


Figure 2: This figure illustrates the hierarchical structure of representation learning techniques for ECG analysis, encompassing deep learning techniques, self-supervised learning approaches, and transfer learning with domain adaptation. Each category is further divided into specific methods, architectures, and applications highlighting advancements in ECG feature extraction, classification accuracy, and diagnostic performance.

4 Representation Learning Techniques

4.1 Deep Learning Techniques

Deep learning has revolutionized ECG analysis by advancing feature extraction and classification. Convolutional Neural Networks (CNNs) are particularly effective in processing noisy ECG signals, enhancing rhythm classification accuracy. The Abductive Interpretation for ECG Classification (AIEC) method exemplifies the use of CNN-derived features for arrhythmia classification [18]. Ullah et al. developed a hybrid deep CNN model with a two-step process, using a 1D CNN for initial classification followed by a 2D CNN for image processing, which significantly boosts accuracy [1]. VizECGNet further integrates 1D and 2D CNNs with attention mechanisms for disease classification from ECG images, demonstrating CNNs' versatility [8].

Method Name	Model Architecture	Feature Extraction	Application Scenarios
AIEC[18]	Sequence Classifier	Abductive Interpretation	Arrhythmia Classification
HDCNN[1]	Hybrid Deep Cnn	1D And 2D	Arrhythmia Classification
VECG[8]	Multi-modal Architecture	Attention Mechanisms	Disease Classification
ISEnet[19]	Isenet Architecture	Attention Mechanism	Arrhythmia Classification
RTECG-CE[16]	Cnn Model	Cnn Model	Disease Diagnosis
GMECC[20]	Seq2seq Architecture	Dense Interpolation	Disease Prediction
PWML[11]	Bi-LSTM	Piece-wise Matching	Real-time Monitoring
MCL[12]	Inceptiontime	Contrastive Learning	Arrhythmia Classification
ECGMatch[13]	Semi-supervised Learning	Data Augmentation	Disease Prediction

Table 1: Overview of various deep learning methods applied to ECG analysis, detailing their model architectures, feature extraction techniques, and application scenarios. The table highlights the diversity in approaches, including CNNs, attention mechanisms, and multimodal learning frameworks, underscoring their roles in arrhythmia classification, disease diagnosis, and real-time monitoring.

As illustrated in Figure 3, the categorization of deep learning techniques applied in ECG analysis highlights CNN-based models, attention mechanisms, and generative and contrastive models as key approaches. Attention mechanisms enhance CNN capabilities, as in the Information-Based Attention Convolutional Neural Network (ISEnet), which improves feature extraction from multi-lead ECG data [19]. Chaotic encryption combined with CNNs addresses security and real-time disease detection [16]. Generative modeling, such as Seq2Seq architecture, enables latent representation generation for accurate classification with limited channels [20]. The piece-wise matching layer technique further enhances performance by extracting morphological and statistical features [11].

Multimodal Contrastive Learning (MCL) aligns embeddings of incomplete and complete ECG signals, improving classification [12]. ECGMatch, a multi-label semi-supervised framework, predicts multiple diseases from limited labeled samples [13]. Various deep learning architectures, including ResNet, Inception, and LSTM networks, demonstrate robust performance, underscoring deep learning's transformative impact on ECG analysis [14]. Table 1 provides a comprehensive overview of the deep learning techniques employed in ECG analysis, emphasizing the diverse model architectures and their specific applications in the field.

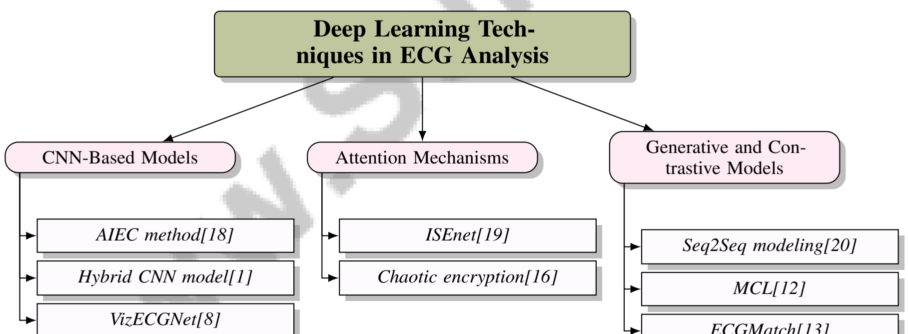


Figure 3: This figure illustrates the categorization of deep learning techniques applied in ECG analysis, highlighting CNN-based models, attention mechanisms, and generative and contrastive models as key approaches.

4.2 Self-Supervised Learning Approaches

Self-supervised learning addresses the limitations of traditional supervised learning in ECG analysis by leveraging vast unlabeled data to develop models that learn meaningful representations. This approach enhances performance in ECG classification and disease prediction, achieving results comparable to supervised methods while improving label efficiency and noise robustness. Techniques like contrastive learning and masked autoencoders facilitate discriminative feature extraction with minimal labeled data [23, 51, 52, 53].

As illustrated in Figure 4, the hierarchical structure of self-supervised learning techniques, applications, and challenges in ECG analysis highlights key methods and their roles in advancing ECG classification, disease prediction, and report generation. The ECG Pre-training Method (EPM) emphasizes lead-agnostic learning, capturing local and global dependencies to improve performance with limited labeled data [54]. Contrastive self-supervised learning highlights the importance of augmentation types and intensities, with augmentations like Gaussian noise yielding optimal outcomes [12, 34, 55]. Integrating clinical knowledge into self-supervised frameworks significantly improves ECG classification, leveraging Large Language Models (LLMs) for accurate abnormality identification and report generation [56, 52].

Generative models, especially -Variational Autoencoders (-VAEs), effectively extract features from ECG data, enhancing predictive accuracy. Novel VAE variants maintain signal fidelity while simplifying data for prediction tasks, matching CNN performance with less data [57, 58, 59]. Self-supervised learning also enhances automated report generation systems, producing more informative ECG analyses [34]. This approach advances diagnostic accuracy without extensive data augmentation, crucial for addressing ECG dataset challenges [21].

Self-supervised learning models align representations of incomplete and complete ECG signals, improving diagnostic performance even with fewer leads [12]. These advancements promise enhanced accuracy, efficiency, and interpretability in cardiovascular diagnostics, leading to better patient outcomes.

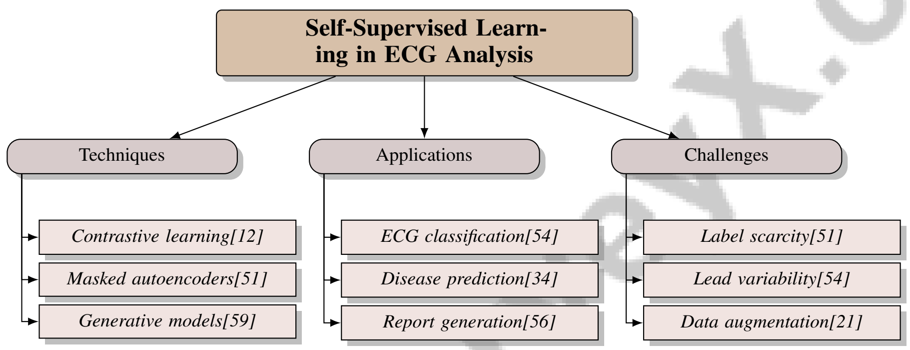


Figure 4: This figure illustrates the hierarchical structure of self-supervised learning techniques, applications, and challenges in ECG analysis, highlighting key methods and their roles in advancing ECG classification, disease prediction, and report generation.

4.3 Transfer Learning and Domain Adaptation

Transfer learning and domain adaptation enhance ECG data analysis by applying models trained in one domain to new domains, improving diagnostic accuracy and efficiency [8]. Leveraging pre-trained models allows adaptation from large datasets to smaller, domain-specific ones, beneficial in medical contexts with scarce labeled data [15]. CNN architectures with dynamic receptive field mechanisms improve classification accuracy by learning from complex ECG data without prior arrhythmia location knowledge [11].

Transfer learning with ensemble architectures like InceptionResNetV2 enhances feature extraction and classification, maintaining high performance with limited data [15]. VizECGNet exemplifies the potential of combining image and signal data through knowledge distillation, ensuring high accuracy even with image data alone [8]. Domain adaptation techniques address ECG signal morphology variability, enhancing diagnostic reliability and accuracy through cross-modal representation learning and advanced methods like optimal transport for knowledge alignment [56, 34, 52, 33]. Dual-attention mechanisms and hierarchical transformer architectures improve model adaptation to new domains, boosting diagnostic performance across clinical scenarios.

The development of transfer learning and domain adaptation techniques, supported by large-scale clinical datasets, significantly advances ECG analysis precision and interpretability. These methodologies leverage advanced machine learning and multimodal data integration to enhance cardiovascular health outcomes by facilitating precise diagnostics across healthcare environments. Models like CardiacNets utilize cross-modal learning to enhance ECG analysis and generate high-quality cardiac images, improving diagnostic accuracy. Integrating real-time medical record access and smart healthcare frameworks further enhances early disease detection and intervention, reducing severe cardiovascular event risks [23, 25, 56, 60].

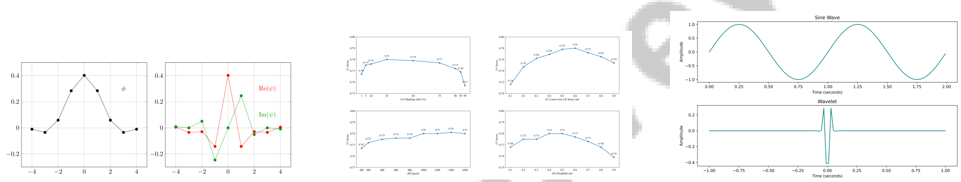
5 Signal Processing and Feature Extraction

5.1 Transformation and Preprocessing Techniques

Preprocessing is crucial in ECG analysis for improving signal quality and ensuring accurate cardiovascular diagnostics. ECG signals often suffer from noise such as baseline wander, power line interference, and muscle noise, which can hinder interpretation [1]. Techniques like the Median Filter and Discrete Wavelet Transform (DWT) are employed to reduce noise while preserving cardiac information [1]. Beyond noise reduction, preprocessing involves extracting cardiac cycles and comprehensive feature sets. For instance, the Seq2Seq model captures essential signal features from limited channel data, facilitating classification [20]. Feature selection methods, including LASSO and Chi-Square, optimize these sets for enhanced classification outcomes.

Advanced preprocessing integrates security measures, such as chaotic encryption, to securely transmit ECG signals while diagnosing heart abnormalities [16]. This dual approach ensures data integrity in real-time applications. Incorporating artificial ECG models simulating specific health conditions enhances classifier performance by broadening ECG data understanding across various health contexts. Using synthetic datasets alongside real data improves deep neural network training, addressing the challenge of limited labeled data and enabling models to capture intricate ECG patterns [61, 33].

Large, annotated ECG datasets, conforming to standards like SCP-ECG, are vital for advancing machine learning models, providing a robust foundation for training and validation [32, 27]. Techniques such as adaptive filtering and DWT enhance signal quality by minimizing noise, facilitating reliable feature extraction and classification in real-time applications.



(a) Graph of a complex function's real and imaginary parts [62]

(b) Machine Learning Model Performance [63]

(c) Sine Wave and Wavelet [64]

Figure 5: Examples of Transformation and Preprocessing Techniques

Figure 5 illustrates the diverse techniques in data transformation and preprocessing. The first image shows a graph of a complex function's components, highlighting data representation intricacies. The second image presents model performance metrics, emphasizing parameter settings' influence on efficacy. The comparison between a sine wave and a wavelet underscores wavelets' utility in capturing sharp transitions and localized features, surpassing simpler periodic functions [62, 63, 64].

5.2 Advanced Feature Extraction Methods

Advanced feature extraction methods are essential for improving ECG analysis accuracy and interpretability, offering insights into cardiovascular health. Techniques like Fuzzy Logic, ANN, Genetic Algorithms, and SVM enhance diagnostic accuracy by extracting significant features from ECG signals [29, 32]. Spectral analysis, such as FFT, decomposes ECG signals into frequency components, identifying subtle characteristics crucial for differentiating cardiac conditions [65].

As illustrated in Figure 6, the hierarchical categorization of advanced feature extraction methods in ECG analysis highlights key techniques, innovative methods, and deep learning models that enhance diagnostic accuracy and efficiency. Innovative methods like dual-scale representation and lead-orthogonal attention mechanisms, as seen in the DLTM-ECG framework, enhance feature extraction by analyzing signals across scales, improving classification accuracy [66]. Deep learning models, such as AlexNet, effectively extract features from fused ECG images, improving analysis accuracy and efficiency [67]. Generative models like MEGAN enhance feature extraction by learning complex patterns, improving classification even with limited channels [42].

Extracting autoregressive coefficients and statistical parameters from ECG signals enhances classification performance, providing a comprehensive framework that generalizes across diverse patient data [68]. The Pan Tompkins algorithm, which extracts fifteen essential features, establishes a solid foundation for accurate diagnostics [50]. Integrating CNN-based architectures with cross-attention mechanisms enhances multimodal integration efficacy, addressing traditional ECG analysis challenges and supporting more accurate health assessments [43].

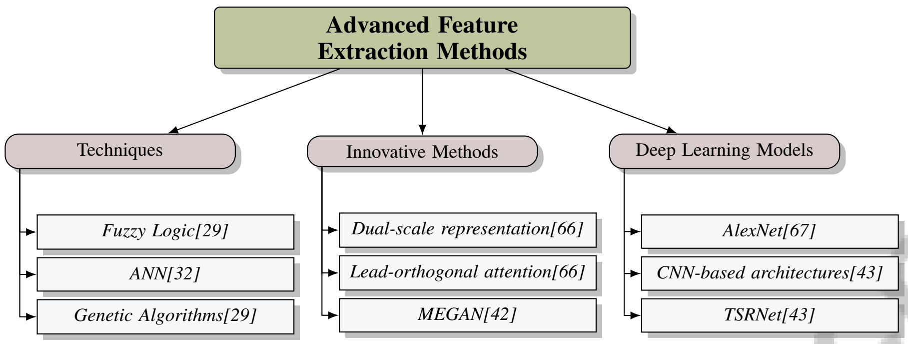


Figure 6: This figure illustrates the hierarchical categorization of advanced feature extraction methods in ECG analysis, highlighting key techniques, innovative methods, and deep learning models that enhance diagnostic accuracy and efficiency.

5.3 Augmentation and Noise Handling Strategies

Augmentation and noise handling are crucial in ECG preprocessing, enhancing diagnostic models' robustness and accuracy. Strategies like TaskAug optimize performance on task-specific ECG problems by adapting to unique characteristics. Advanced noise detection frameworks use deep learning to identify and mitigate noisy artifacts, improving diagnostic reliability [55, 34]. Techniques like Gaussian noise introduction simulate real-world variability, enhancing model generalization [55].

Augmentation methods, including scaling, permutation, and time warping, expand datasets, enabling models to learn invariant features [55]. Vertical and horizontal flips, along with zero masking, introduce variability, improving model resilience to signal presentation variations [55]. Effective noise handling strategies, including adaptive filters and DWT, eliminate unwanted artifacts, enhancing signal quality and ensuring accurate feature extraction. These strategies are vital for ambulatory monitoring systems, aiding in early disease detection and real-time analysis in portable devices [69, 29].

Integrating augmentation and noise handling in ECG processing is crucial for developing robust diagnostic models. These approaches enhance model generalization, accuracy, and consistency across clinical settings. Techniques like knowledge transfer from LLMs to ECG and multimodal data integration facilitate high-quality disease detection and automatic report generation. Novel loss functions and contrastive learning strategies ensure effective inter-modal relationship alignment, promoting robust performance in medical diagnostics [56, 38].

6 Applications in Cardiovascular Health

Exploring multimodal representation learning in cardiovascular health reveals its transformative potential in diagnostics, enhancing traditional methods' accuracy and fostering innovative monitoring approaches. A pivotal application is in arrhythmia detection and monitoring, where these techniques integrate multimodal data to address cardiac irregularities, significantly improving patient outcomes.

6.1 Arrhythmia Detection and Monitoring

Multimodal representation learning has significantly advanced arrhythmia detection and monitoring, overcoming limitations of traditional ECG interpretation reliant on inconsistent manual analysis, vulnerable to errors in diagnosing complex arrhythmias like atrial fibrillation [70]. By integrating diverse data modalities and advanced learning frameworks, these methods enhance diagnostic accuracy and interpretability. Hybrid CNN architectures have been crucial, with models like a hybrid deep CNN achieving 97.38% accuracy for 1D CNN and 99.02% for 2D CNN in arrhythmia detection [1].

Models like ISE-net, utilizing multiple lead ECG signals, enhance classification accuracy through comprehensive ECG data utilization [19]. VizECGNet, classifying diseases from printed ECG images, exemplifies multimodal approaches' effectiveness in distinguishing similar cardiac conditions [8, 50]. In wearable applications, these methodologies improve accuracy and noise resistance, suitable for continuous monitoring and early anomaly detection [65]. Systems like ECG-Chat enhance cardiac disease diagnosis through ECG report generation and multimodal dialogues [71]. Secure telemedicine and remote monitoring systems, ensuring high data security without compromising real-time performance, are essential in resource-constrained environments [16]. Multimodal representation learning in ECG analysis promises enhanced cardiovascular diagnostics, leveraging diverse data modalities and sophisticated frameworks to revolutionize health monitoring and treatment, optimizing arrhythmia detection and monitoring [17].

6.2 Myocardial Infarction and Heart Disease Diagnosis

Multimodal representation learning significantly advances myocardial infarction (MI) and heart disease diagnosis by integrating ECG signals, clinical text, and imaging data. Traditional ECG methods struggle with MI diagnosis due to waveform complexities affected by noise and artifacts, requiring meticulous preprocessing for accuracy. Challenges in distinguishing normal and abnormal rhythms, compounded by noise, hinder diagnostic processes, potentially leading to misdiagnoses. Recent deep learning advancements aim to improve noisy ECG sample detection, yet clinical integration remains limited, underscoring the need for further research [69, 32, 72].

To illustrate the hierarchical structure of multimodal representation learning in cardiovascular diagnosis, Figure 7 presents a visual summary that focuses on ECG analysis challenges, advanced learning techniques, and the benefits of data integration. This figure underscores how multimodal representation learning addresses these challenges by leveraging complementary information from multiple sources, enhancing diagnostic accuracy and robustness. CNN-based architectures have improved ECG classification, with hybrid 1D and 2D CNNs capturing local and global features for enhanced performance. These models, combined with attention mechanisms, improve generalization across diverse clinical scenarios [13].

Integrating ECG and cardiac magnetic resonance imaging significantly enhances MI and heart disease diagnostic accuracy [12], providing comprehensive cardiac function assessments and improving diagnostic model robustness [25]. Self-supervised learning approaches enhance ECG analysis capabilities, improving model generalization with limited labeled data. These representation learning advancements promise significant improvements in MI and heart disease diagnosis, leading to better patient outcomes and cardiovascular management. Large-scale cross-modality pretrained models like CardiacNets exemplify advanced machine learning's potential in enhancing ECG analysis by integrating clinical text with ECG signals, crucial for distinguishing normal aging effects from pathological changes in short-duration 12-lead ECGs [25, 26]. CNN-based models using cross-attention mechanisms further enhance ECG analysis [43].

Multimodal data integration, as demonstrated with 2.3 million ECG recordings, advances health analysis [41]. Integrating data sources like cardiac magnetic resonance imaging and clinical text reports improves diagnostic performance, offering comprehensive cardiac health assessments [25]. Enhancing ECG-language question-answering tasks' generalization performance, multimodal representation learning frameworks improve diagnostic accuracy and patient outcomes [12].

6.3 Personalized and Continuous Monitoring

Integrating multimodal representation learning into ECG analysis has advanced personalized and continuous cardiovascular monitoring. By combining ECG signals with clinical reports, patient metadata, and textual descriptions, this approach develops robust, generalizable models for diverse clinical settings, including under-resourced regions [73, 34, 12, 47, 24]. Leveraging multiple physiological signals and data modalities provides comprehensive cardiovascular health assessments, enhancing prolonged monitoring. Practical applications, like semi-supervised learning in ECG signal classification with limited labeled data, are crucial in real-world scenarios [13]. Large-scale datasets with predefined train-test splits enhance model reliability and facilitate clinical translation [14]. Integrating clinical text and ECG signals improves diagnostic accuracy, enabling targeted treatment strategies [25]. Multimodal representation learning supports personalized and continuous monitoring

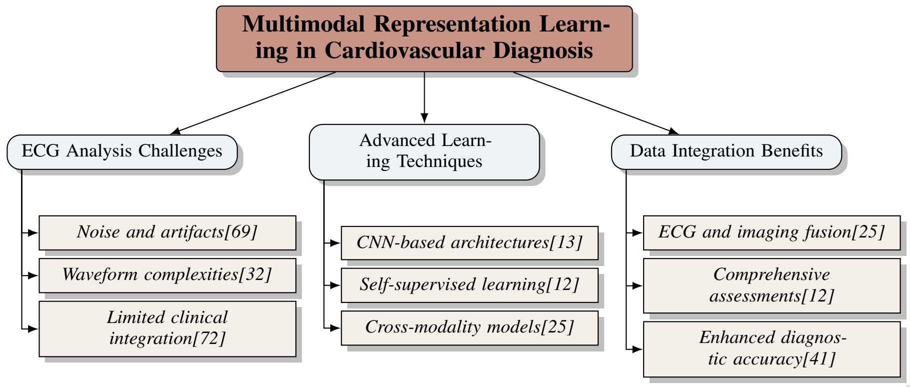


Figure 7: This figure illustrates the hierarchical structure of multimodal representation learning in cardiovascular diagnosis, focusing on ECG analysis challenges, advanced learning techniques, and the benefits of data integration.

systems, providing real-time insights for timely interventions and personalized treatment plans. These techniques enhance cardiovascular health assessments' precision, fostering targeted and effective treatment strategies [23]. Extending beyond traditional clinical environments, these methods hold potential in resource-limited settings, offering comprehensive diagnostic insights and improving patient outcomes [9]. By leveraging multiple data sources' strengths, these techniques provide a comprehensive understanding of cardiovascular health, leading to improved patient outcomes and efficient healthcare delivery."

7 Conclusion

Advancements in multimodal representation learning for electrocardiogram (ECG) analysis underscore the pivotal role of signal processing and feature extraction in enhancing cardiovascular diagnostics. This approach integrates diverse data modalities, such as ECG signals, phonocardiograms (PCG), and clinical text, providing a comprehensive view of cardiac function. Generative models like Auto-TTE synthesize 12-lead ECGs from clinical text, facilitating automated decision support and classification, thereby enhancing diagnostic insights [33, 35]. By leveraging deep learning, self-supervised, and transfer learning, these methods address traditional ECG analysis limitations, which often rely on manual interpretation and extensive labeled datasets.

Integrating multimodal data sources enhances diagnostic accuracy and interpretability, leading to improved outcomes and personalized treatment strategies [41]. The survey highlights the transformative potential of future advancements in multimodal representation learning, particularly in developing efficient ECG retrieval systems using large language models and vision-language paradigms. These innovations promise to enhance clinical decision support and diagnostic service accessibility, especially in underdeveloped regions [12, 24, 35]. Continued research is crucial for improving cardiovascular diagnostics' precision, efficiency, and accessibility, ultimately benefiting patient outcomes and healthcare delivery.

7.1 Signal Processing and Feature Extraction

ECG data analysis for cardiovascular monitoring relies on advanced signal processing and feature extraction techniques essential for accurately interpreting the heart's electrical activity. These waveforms are often compromised by noise, making robust preprocessing, feature extraction, and classification techniques vital for reliable diagnostics [69, 6, 74, 32]. Signal processing techniques, such as wavelet transforms and adaptive filtering, are crucial for noise mitigation and signal quality enhancement, paving the way for precise feature extraction [29, 32].

Feature extraction from ECG signals involves identifying characteristics indicative of cardiovascular health. Techniques like Discrete Wavelet Transform (DWT) decompose ECG signals into frequency bands, facilitating the extraction of essential time-frequency features [37]. Machine learning models, particularly CNNs, enhance automatic feature extraction, improving diagnostic accuracy and effi-

AI-generated, for reference only.

ciency. Advanced methods, such as the DLTM-ECG framework, leverage dual-scale representation and lead-orthogonal attention mechanisms to capture complex wave segments, critical for accurate diagnosis [66]. Integrating multimodal data, including ECG signals and clinical text, further enhances diagnostic accuracy and interpretability [75, 38].

Sophisticated algorithms, particularly CNNs and generative models, have markedly improved ECG interpretation accuracy, demonstrated by advancements in automated report generation and ECG synthesis from clinical texts [33, 45, 52, 76, 35]. Techniques combining Fast Fourier Transform (FFT) with RR interval features significantly improve ECG classification accuracy [65]. These methods provide a robust framework for ECG signal analysis, enabling subtle pattern identification crucial for accurate diagnostics.

Integrating multimodal data, including ECG and PCG signals, enhances ECG analysis robustness and accuracy by offering a comprehensive view of cardiovascular health [77]. This approach allows cross-verification from multiple modalities, improving diagnostic models' reliability and generalization [38]. Continued development of advanced signal processing and feature extraction techniques is essential for overcoming traditional ECG analysis limitations, enhancing heart disease detection and prognosis accuracy [29, 32, 78].

7.2 Enhancements in Diagnostic Tools

The integration of multimodal representation learning into ECG analysis has significantly enhanced diagnostic tools, improving accuracy, efficiency, and interpretability in cardiovascular assessments. These advancements address traditional ECG interpretation limitations, which often rely on manual analysis prone to human error [10].

By integrating information from multiple sources, such as clinical text and imaging data, these techniques enhance diagnostic accuracy and reliability, facilitating effective disease detection and monitoring [75, 12]. Advanced machine learning models, especially deep CNNs, have significantly improved ECG signal classification accuracy. Hybrid architectures combining 1D and 2D CNNs enhance arrhythmia detection reliability, making them ideal for real-time wearable health monitoring [79, 80, 1, 65, 17].

Generative models, like the MEGAN framework, improve ECG classification by capturing complex patterns and enhancing diagnostic robustness, particularly in limited channel availability scenarios. Integrating multimodal data, including ECG signals, PCG signals, and clinical text, has further enhanced diagnostic accuracy and interpretability, improving cardiovascular health understanding through advanced machine learning techniques [44, 24]. Utilizing large-scale annotated datasets, such as those conforming to the SCP-ECG standard, has been instrumental in advancing machine learning models for ECG analysis, ensuring reliability and accuracy [76, 35].

Ongoing development of sophisticated algorithms and data integration techniques is crucial for significantly improving ECG analysis precision and practical application in clinical settings, enhancing preprocessing, feature extraction, and classification processes to improve diagnostic accuracy for various cardiac and non-cardiac conditions [76, 35, 81].

7.3 Applications in Resource-Limited Settings

The deployment of ECG multimodal representation learning in resource-limited settings holds significant potential for enhancing cardiovascular diagnostics, particularly where specialized medical equipment and personnel are scarce. Traditional 12-lead ECG systems may be unavailable, necessitating innovative solutions that effectively operate in low-resource environments [12].

A primary challenge is the scarcity of labeled data for training robust machine learning models. Semi-supervised learning methods, such as ECGMatch, leverage unlabeled data to improve model performance, enabling accurate ECG classification even with limited labeled access [13]. Generative modeling approaches, like -Variational Autoencoders (-VAEs), enhance ECG classification performance in scenarios with limited channel availability, demonstrating versatility across diverse diagnostic tasks [20].

Integrating multimodal data sources, such as ECG signals and clinical text, improves diagnostic performance in resource-limited settings, allowing healthcare practitioners to make informed decisions even without complete diagnostic equipment [75]. Developing secure, efficient systems for ECG data transmission and analysis is critical in such settings, where data security and efficiency are paramount for telemedicine and remote monitoring [16].

8 Challenges and Future Directions

8.1 Data Limitations and Imbalance

ECG data analysis faces significant challenges due to data limitations and class imbalance, which hinder the accuracy and generalizability of machine learning models, especially in resource-limited settings with restricted access to large, annotated datasets [13]. The scarcity of labeled ECG data, crucial for robust model training, is exacerbated by the labor-intensive nature of manual annotation and the imbalanced distribution of cardiac abnormalities, resulting in models biased towards majority classes and suboptimal performance on minority classes [14]. Developing large, annotated datasets like the SCP-ECG standard, along with resources such as the ECG-Image-Database and the ECG-Text Pre-training (ETP) framework, is essential for addressing these challenges by enhancing cross-modal representation learning and ensuring reliable ECG interpretations [34, 36]. Semi-supervised learning methods, such as ECGMatch, utilize unlabeled data to improve model generalization across diverse datasets [13], while integrating multimodal data sources, including ECG signals and clinical text reports, further enhances diagnostic performance [75].

8.2 Model Generalization and Overfitting

Achieving model generalization while avoiding overfitting is a critical challenge in ECG data analysis using machine learning. Overfitting occurs when a model performs well on training data but poorly on unseen data, compromising diagnostic performance in real-world scenarios [13]. The complexity of deep learning models, like CNNs, contributes to this issue due to their high capacity and flexibility, which can lead to overfitting, particularly with limited labeled ECG data. To mitigate overfitting, regularization techniques (e.g., dropout, weight decay) are used to prevent reliance on specific features [11], while data augmentation methods, such as scaling and time warping, expand the dataset to enhance generalization [55]. Integrating multimodal data sources, like ECG and PCG signals, provides a more comprehensive view of cardiovascular health, improving generalization [77]. Self-supervised learning techniques, such as the ECG Pre-training Method (EPM), leverage large amounts of unlabeled data for robust ECG signal representation, enhancing model generalization [54]. Addressing these issues is crucial for advancing ECG analysis precision and applicability in clinical settings [25, 23, 12, 24].

8.3 Computational Complexity and Efficiency

Advanced ECG analysis techniques, particularly those involving multimodal representation learning, present substantial computational challenges that must be addressed for practical clinical application. These models often require increased computational resources, hindering deployment in resource-constrained environments [43]. High computational costs associated with training deep learning models, especially those utilizing complex architectures like CNNs and GANs, pose barriers to implementation [16]. Strategies to enhance computational efficiency include utilizing lightweight neural network architectures, such as MobileNet, which reduce complexity while maintaining high classification accuracy, making them suitable for real-time applications [43]. Chaotic encryption techniques offer solutions for enhancing data security and computational efficiency, particularly in resource-constrained settings [16]. Lightweight models based on one-dimensional CNNs facilitate real-time monitoring through efficient feature extraction methods [1]. Integrating multimodal data sources, such as ECG signals and clinical text reports, optimizes model performance and enhances ECG analysis efficiency across diverse clinical settings [75].

8.4 Integration with Clinical Practice

Integrating multimodal representation learning into clinical practice enhances cardiovascular diagnostics' accuracy, efficiency, and interpretability, but presents challenges that must be addressed [13]. A primary challenge is the computational demands of advanced machine learning models,

AI-generated, for reference only.

such as deep CNNs and generative models, which require substantial resources for training and deployment, posing barriers in resource-constrained clinical settings [16, 43]. Utilizing lightweight architectures like MobileNet is crucial for enabling real-time ECG analysis and diagnostics. The availability of comprehensive, well-annotated datasets that reflect diverse patient populations is essential for training robust models capable of generalizing across various clinical scenarios [14, 9]. Integrating multimodal data sources, including ECG signals, PCG signals, and clinical text reports, enhances diagnostic accuracy and interpretability [75, 38]. Self-supervised learning techniques, such as the ECG Pre-training Method (EPM), improve model generalization by leveraging unlabeled data, which is particularly valuable in resource-limited settings [54, 13]. Addressing standardization and validation issues is essential for effectively implementing multimodal representation learning in clinical practice, requiring comprehensive evaluation frameworks.

9 Conclusion

The incorporation of multimodal representation learning into ECG analysis markedly advances cardiovascular diagnostics by integrating diverse data modalities, such as ECG signals, PCG, and clinical text reports. This multifaceted approach enhances understanding of cardiac function through sophisticated methodologies, including autoregressive generative models conditioned on clinical text for synthesizing comprehensive 12-lead ECGs, automated clinician-like interpretations of ECG data, and multimodal techniques aligning ECG signals with diagnostic reports. These strategies significantly improve diagnostic accuracy and support automated clinical decision-making, proving invaluable in both advanced research and practical healthcare applications [24, 33, 35, 32]. By leveraging advanced machine learning techniques, such as deep learning, self-supervised learning, and transfer learning, multimodal representation learning addresses the limitations of traditional ECG analysis, which often depends on manual interpretation and extensive labeled data.

The integration of multimodal data sources, including ECG signals, CMR images, and clinical text, enhances diagnostic accuracy and interpretability in cardiovascular assessments. This comprehensive approach allows healthcare practitioners to gain a holistic understanding of a patient's cardiovascular health, leading to improved diagnostic outcomes and personalized treatment strategies [41].

The survey underscores the pivotal role of multimodal representation learning in ECG analysis, highlighting innovative approaches like contrastive learning and vision-language models that enhance diagnostic accuracy with fewer ECG leads, thereby addressing accessibility issues in clinical settings. Future advancements in this domain have the potential to revolutionize ECG interpretation and retrieval systems, enhancing their efficiency and practicality, particularly in underdeveloped regions [12, 24, 73].

9.1 Enhancements in Diagnostic Tools

The integration of multimodal ECG analysis into diagnostic tools has significantly improved the accuracy, efficiency, and interpretability of cardiovascular diagnostics. By leveraging diverse data modalities, such as ECG images, clinical text, and physiological signals, these advancements address the limitations of traditional ECG interpretation methods, which are often prone to human error. Deep learning models, such as those developed by Ribeiro et al., offer competitive alternatives to classical ECG classification methods, demonstrating high accuracy in the automatic diagnosis of short-duration 12-lead ECGs [82].

Innovative approaches have been crucial in enhancing diagnostic accuracy. Dua et al.'s two-stage framework effectively detects noisy ECG samples, outperforming existing methods and showcasing its potential for real-world applications in ECG signal analysis and diagnosis [69]. Similarly, Vulaj et al.'s method achieved 100

The automated digitization of ECGs, combined with machine learning, enables robust analysis, paving the way for advancements in computational diagnostics [83]. This approach enhances cardiovascular diagnostic precision and streamlines ECG data integration into clinical workflows, improving overall healthcare delivery efficiency.

Moreover, multimodal learning frameworks have significantly improved diagnostic accuracy across multiple benchmarks. These frameworks utilize the strengths of various data modalities, including ECG signals and clinical text reports, to deliver holistic evaluations of cardiovascular health. By integrating advanced machine learning techniques such as contrastive learning and vision-language alignment, these approaches facilitate the automatic identification of similar clinical cases and the generation of clinician-like interpretations of ECG data. This comprehensive methodology enhances diagnostic accuracy for various cardiovascular diseases, such as coronary artery disease and cardiomyopathy, supporting informed clinical decision-making, particularly in resource-limited settings [12, 24, 35, 25]. Self-supervised anomaly detection methods have further contributed to significant improvements in diagnostic accuracy for both common and rare cardiac conditions.

The integration of multimodal representation learning into ECG analysis has led to notable enhancements in diagnostic tools, improving accuracy, efficiency, and interpretability in cardiovascular diagnostics. By incorporating diverse data modalities such as CMR images, ECG signals, and clinical metadata with advanced learning frameworks like self-supervised and contrastive learning, recent innovations are set to transform cardiovascular health monitoring and treatment. These approaches enable comprehensive insights into patients' cardiovascular conditions, even from limited annotated datasets, significantly improving disease prediction accuracy and enhancing diagnostic efficiency through automated ECG interpretation using vision-language models. Ultimately, these advancements are expected to improve patient outcomes and streamline healthcare delivery, particularly in resource-limited settings [23, 12, 24].

9.2 Applications in Resource-Limited Settings

The application of ECG multimodal representation learning in resource-limited settings presents transformative potential for advancing cardiovascular health diagnostics, particularly where access to specialized medical equipment and personnel is restricted. Traditional 12-lead ECG systems are often unavailable in such environments, necessitating innovative solutions that can operate effectively with fewer leads while maintaining diagnostic accuracy [12].

A significant challenge in these settings is the scarcity of labeled data, crucial for training effective machine learning models. Semi-supervised learning approaches, like ECGMatch, leverage abundant unlabeled data to enhance model performance, even when labeled data is limited [13]. These methods enable accurate ECG classification despite limited labeled data availability.

The LightX3ECG system exemplifies a lightweight and explainable deep learning approach that addresses the need for accessible ECG diagnostics in resource-constrained environments. By optimizing model architecture for efficiency and interpretability, LightX3ECG ensures accurate and reliable ECG analysis with minimal computational resources [84]. This system underscores the potential of advanced machine learning techniques to enhance ECG analysis and diagnostic accuracy, particularly in settings with limited access to sophisticated medical equipment [84].

Furthermore, integrating multimodal data sources, such as ECG signals and clinical text reports, has proven beneficial in resource-limited environments by providing comprehensive assessments of cardiovascular health [75]. This approach enables healthcare practitioners to make informed decisions, improving patient outcomes and optimizing healthcare delivery even in the absence of complete diagnostic equipment.

The application of ECG data for liver disease detection illustrates the scalability and affordability of ECG-based diagnostics in resource-limited settings [5]. This highlights the versatility of ECG as a diagnostic tool, capable of addressing a wide range of health conditions beyond traditional cardiac diseases.

Developing secure and efficient systems for ECG data transmission and analysis is crucial in resource-limited settings, where real-time processing capabilities are essential for immediate medical responses. The application of chaotic encryption techniques in ECG analysis has shown potential in enhancing data security and efficiency, making it a viable option for telemedicine and remote monitoring in low-resource environments [16]. These advancements in secure and efficient ECG analysis systems underscore the potential of multimodal representation learning to revolutionize cardiovascular health diagnostics in resource-constrained environments.

10 Challenges and Future Directions

10.1 Data Limitations and Imbalance

ECG data analysis for cardiovascular diagnostics is hampered by significant challenges, notably data limitations and class imbalance, which undermine the efficacy and generalizability of machine learning models. The scarcity of labeled ECG data, compounded by labor-intensive annotation and uneven cardiac abnormality distribution, biases models toward majority classes, leading to suboptimal minority class performance [13, 11]. Addressing these issues requires comprehensive annotated datasets like PTB-XL and CPSC2018, which enhance model training and validation by incorporating real-world artifacts and employing advanced techniques such as ECG-Text Pre-training and retrieval-augmented self-supervised learning [33, 36, 34, 52, 35].

Semi-supervised learning methods, such as ECGMatch, leverage unlabeled data to improve model performance and generalization across datasets [13]. Integrating multimodal data, including ECG signals and clinical text, enhances diagnostic accuracy by offering a comprehensive cardiovascular health assessment [75]. Data augmentation techniques like Gaussian noise, scaling, permutation, and time warping expand training data and bolster model robustness, addressing data limitations and class imbalance [55, 58]. Nonetheless, domain generalization approaches face performance drops on specific classes due to ECG data complexities and class imbalance [85, 58]. In resource-limited settings, the challenge of missing leads necessitates innovative solutions to maintain diagnostic accuracy with fewer leads [12].

10.2 Model Generalization and Overfitting

In ECG analysis, model generalization and overfitting are critical challenges. Overfitting occurs when models excessively learn training data, capturing noise as patterns and performing poorly on new data [13]. This issue is exacerbated by ECG signal complexity and variability, resulting in models that excel in training but fail to generalize [37]. The complexity of deep learning architectures, like CNNs, increases overfitting risks, especially given the scarcity of large labeled datasets in medical domains [43, 13]. Regularization techniques, such as dropout and weight decay, prevent excessive model complexity, while data augmentation methods enhance generalization by exposing models to varied conditions.

Multimodal data integration, encompassing ECG, PCG signals, and clinical text, improves generalization by providing a comprehensive cardiovascular health view [23, 38]. Self-supervised learning techniques, like ECG Pre-training Method (EPM), utilize large unlabeled data to learn robust ECG representations, enhancing generalization across diverse datasets [54, 13]. Transfer learning and domain adaptation techniques are crucial for improving model generalization and mitigating overfitting. These methods address annotated dataset scarcity and model adaptation challenges to out-of-distribution data, crucial in clinical settings with variable ECG data. For instance, the ECG-Text Pre-training (ETP) framework enhances zero-shot classification tasks by linking ECG signals with textual reports. Empirical studies show transfer learning's effectiveness for small datasets, while domain generalization approaches maintain performance across diverse data distributions [85, 86, 34]. However, reliance on specific datasets may still challenge generalizability, potentially leading to performance drops on other ECG signal types.

Addressing model generalization and overfitting challenges is vital for advancing ECG analysis precision and applicability in clinical settings. By leveraging advanced techniques like semi-supervised learning, multimodal data integration, and self-supervised learning, researchers enhance cardiovascular diagnostics' robustness and accuracy. Models integrating cardiac magnetic resonance images, ECG signals, and clinical data provide comprehensive cardiovascular health views, improving predictive capabilities even with limited labeled datasets. Techniques like masked autoencoders and multi-modal contrastive learning effectively transfer knowledge across modalities, enhancing performance in diagnosing conditions like myocardial infarction. Moreover, semi-supervised models like ECGMatch address label scarcity and co-occurring cardiovascular diseases, enabling robust predictions with minimal supervision and improving diagnostic accuracy across diverse healthcare settings [23, 12, 24, 13].

AI-generated, for reference only.

10.3 Computational Complexity and Efficiency

The implementation of advanced ECG analysis techniques, particularly multimodal representation learning, poses significant computational challenges that must be addressed for practical clinical applicability. The complexity of these models often results in increased computational demands, hindering deployment in resource-constrained environments with limited processing power and storage [43]. A primary challenge is the high computational cost of training and deploying deep learning models, especially those using complex architectures like CNNs and Transformer-based models, which require substantial computational resources and pose barriers to real-time clinical practice integration [87].

To enhance computational efficiency, various strategies have been proposed. Optimization processes, such as Enhanced Harris Hawks Optimization-Support Vector Machine (EHO-SVM), significantly reduce computational complexity by optimizing feature selection, improving performance and efficiency in ECG analysis. Lightweight neural network architectures, like MobileNet, maintain high diagnostic accuracy while reducing computational demands [43]. Techniques such as stacked denoising autoencoders for noise reduction optimize feature extraction processes, enhancing computational efficiency and real-time monitoring system development [66].

Integrating multimodal data sources, such as ECG signals and clinical text reports, further enhances efficiency by providing a comprehensive cardiovascular diagnostic framework [75]. This approach optimizes model performance, ensuring accurate and efficient diagnostics. Establishing standardized benchmarks and improving model robustness through methodologies like cross-validation and domain generalization are essential for enhancing ECG analysis models' generalizability and reliability across clinical environments. Single-source training methods often yield overly optimistic performance estimates, potentially misleading clinical applications. Techniques like leave-source-out cross-validation provide more accurate performance assessments by accounting for patient population variability. Addressing out-of-distribution (OOD) data challenges is crucial, as models trained on data from one hospital often underperform on data from another. By employing cross-validation and domain generalization, researchers develop methods to capture ECG signals' underlying structures, improving classification accuracy and ensuring models maintain performance across diverse clinical settings [85, 88]. These advancements underscore the need for ongoing efforts to tackle computational complexity and efficiency in ECG analysis, ensuring advanced machine learning models can be effectively integrated into real-world clinical practice.

11 Conclusion

The integration of multimodal representation learning in ECG analysis significantly enhances cardiovascular diagnostics by synthesizing diverse data modalities, such as ECG signals, phonocardiograms (PCG), and clinical text reports. This comprehensive approach fosters a deeper understanding of cardiac function and addresses traditional ECG analysis limitations, which often rely on manual interpretation and extensive labeled datasets. Advanced machine learning techniques, including deep learning, self-supervised learning, and transfer learning, enhance diagnostic accuracy by amalgamating various data sources, including patient metadata, and allow for high performance with fewer leads. Self-supervised learning mitigates label scarcity by extracting valuable representations from ECG data, enhancing efficiency and robustness in downstream tasks without extensive human labeling. Techniques such as contrastive learning and vision-language alignment further optimize diagnostics, improving accessibility, especially in resource-limited settings [12, 24, 73, 51].

Combining multimodal data sources, including ECG signals, cardiac magnetic resonance (CMR) images, and clinical text, improves diagnostic accuracy and interpretability in cardiovascular assessments. This approach enables practitioners to achieve a holistic understanding of a patient's cardiovascular status, facilitating personalized treatment strategies [41].

The survey emphasizes the pivotal role of multimodal representation learning in enhancing cardiovascular diagnostics, highlighting innovative techniques such as contrastive learning and vision-language alignment to improve diagnostic accuracy with limited ECG leads. It discusses future advancements that could revolutionize ECG interpretation and retrieval systems, particularly in resource-limited contexts [12, 24, 73].

11.1 Integration with Clinical Practice

Integrating multimodal representation learning into clinical practice presents challenges and opportunities for enhancing ECG analysis and cardiovascular diagnostics' accuracy and efficiency. A primary challenge is the computational complexity of advanced machine learning models, such as deep CNNs and generative models, which require substantial resources for training and deployment, posing barriers in resource-limited environments. Developing lightweight and efficient models, such as MobileNet, is crucial to reduce computational complexity while maintaining high diagnostic accuracy, facilitating the integration of advanced ECG analysis methods into clinical settings [43].

Comprehensive and well-annotated datasets that reflect diverse patient populations are essential for training robust models capable of generalizing across clinical scenarios. Large-scale annotated datasets, adhering to standards like SCP-ECG, are crucial for this purpose [9]. Future research should focus on creating standardized datasets, enhancing model interpretability, addressing data imbalance, and exploring multimodal data integration to improve ECG analysis [61]. Maintaining diagnostic integrity while employing advanced signal processing techniques is vital [89].

Leveraging multimodal data sources, including ECG and PCG signals, and clinical text reports, offers a promising approach to enhancing cardiovascular diagnostics' accuracy and interpretability in clinical practice [75]. By leveraging multiple data modalities, multimodal representation learning frameworks facilitate comprehensive cardiovascular health assessments, enabling more precise diagnostic decisions [38]. Federated learning and explainable AI introduce opportunities and challenges within clinical workflows, particularly concerning data privacy [90]. Expanding datasets to encompass a wider variety of ECG types and refining models to enhance applicability in clinical environments are critical future research directions [91].

Interdisciplinary collaboration among data scientists, clinicians, and ethicists is essential to address challenges in personalized ECG analysis [92]. This collaboration is crucial for integrating advanced ECG analysis techniques into clinical workflows, enhancing the understanding of cardiovascular aging [26]. A standardized method for evaluating feature importance in ECG analysis aligns algorithmic results with clinical expertise [93].

Future research must focus on developing robust models that effectively identify and exploit temporal variations in ECG data [31]. Incorporating alternative concept discovery techniques can significantly enhance the interpretability of ECG analysis models, facilitating their integration into clinical workflows [31]. This is particularly crucial for automated diagnostic tools, where clinical relevance and interpretability are essential for successful implementation [79].

12 Conclusion

This survey highlights the transformative potential of multimodal representation learning in electrocardiogram (ECG) analysis, significantly enhancing cardiovascular health diagnostics. By integrating diverse data modalities—such as ECG signals, phonocardiograms (PCG), cardiac magnetic resonance (CMR) images, and clinical text reports—this approach provides a comprehensive framework for understanding cardiac function. Advanced machine learning techniques, including deep learning, self-supervised learning, and transfer learning, effectively overcome the limitations of traditional ECG analysis, which often depends on manual interpretation and extensive labeled datasets.

The combination of multimodal data sources with ECG signals markedly improves diagnostic accuracy and interpretability, enabling healthcare practitioners to achieve a holistic understanding of a patient's cardiovascular condition. This integration leads to enhanced diagnostic outcomes and personalized treatment strategies [41]. Furthermore, federated learning techniques have shown promise in real-world arrhythmia classification applications, demonstrating performance comparable to centralized models while reducing execution time [94].

Future advancements in this domain are poised to further enhance the accuracy, efficiency, and robustness of ECG analysis, particularly in developing countries where access to advanced diagnostic tools is limited [95]. Innovative data augmentation techniques and the integration of multimodal data sources are crucial for improving the generalization and robustness of ECG analysis models, ultimately leading to better patient outcomes and more effective healthcare delivery. Future research should focus on refining models, exploring additional data sources, and enhancing interpretability for clinical applications [96].

The proposed T-S reverse detection method effectively learns ECG representations, significantly outperforming baseline methods in atrial fibrillation detection, thus demonstrating the potential of self-supervised learning techniques to enhance model generalization for real-world applications. Additionally, the development of systems for the digitalization of plotted electrocardiograms has shown promising results, underscoring the utility of automated solutions in clinical settings [6]. Moreover, the creation of interpretable ECG beat embedding spaces improves the applicability of machine learning models in clinical environments, facilitating effective ECG classification systems that maintain interpretability while achieving high accuracy.

AI-generated, for reference only.

References

[1] Amin Ullah, Sadaqat ur Rehman, Shanshan Tu, Raja Majid Mehmood, Fawad, and Muhammad Ehatisham-ul Haq. A hybrid deep cnn model for abnormal arrhythmia detection based on cardiac ecg signal. *Sensors*, 21(3):951, 2021.

[2] Geoffrey H. Tison, Jeffrey Zhang, Francesca N. Delling, and Rahul C. Deo. Automated and interpretable patient ecg profiles for disease detection, tracking, and discovery, 2018.

[3] Binhang Yuan and Wenhui Xing. Diagnosing cardiac abnormalities from 12-lead electrocardiograms using enhanced deep convolutional neural networks, 2019.

[4] Ecg beat classification using machine learning and pre-trained convolutional neural networks.

[5] Juan Miguel Lopez Alcaraz, Wilhelm Haverkamp, and Nils Strodthoff. Electrocardiogram-based diagnosis of liver diseases: an externally validated and explainable machine learning approach, 2024.

[6] Manuel Pazos-Santomé, Fernando Martín-Rodríguez, and Mónica Fernández-Barciela. Automated optical reading of scanned ecgs, 2024.

[7] Asim Darwaish, Farid Naït-Abdesselam, and Ashfaq Khokhar. Detection and prediction of cardiac anomalies using wireless body sensors and bayesian belief networks, 2019.

[8] Ju-Hyeon Nam, Seo-Hyung Park, Su Jung Kim, and Sang-Chul Lee. Vizecgnet: Visual ecg image network for cardiovascular diseases classification with multi-modal training and knowledge distillation, 2024.

[9] Mohamed Adel Serhani, Hadeel T. El Kassabi, Heba Ismail, and Alramzana Nujum Navaz. Ecg monitoring systems: Review, architecture, processes, and key challenges. *Sensors*, 20(6):1796, 2020.

[10] Sajad Mousavi, Fatemeh Afghah, and U. Rajendra Acharya. Han-ecg: An interpretable atrial fibrillation detection model using hierarchical attention networks, 2020.

[11] Behzad Ghazanfari, Fatemeh Afghah, and Sixian Zhang. Piece-wise matching layer in representation learning for ecg classification, 2020.

[12] Tue M. Cao, Nhat H. Tran, Phi Le Nguyen, and Hieu Pham. Multimodal contrastive learning for diagnosing cardiovascular diseases from electrocardiography (ecg) signals and patient metadata, 2023.

[13] Rushuang Zhou, Lei Lu, Zijun Liu, Ting Xiang, Zhen Liang, David A. Clifton, Yining Dong, and Yuan-Ting Zhang. Semi-supervised learning for multi-label cardiovascular diseases prediction:a multi-dataset study, 2023.

[14] Nils Strodthoff, Patrick Wagner, Tobias Schaeffter, and Wojciech Samek. Deep learning for ecg analysis: Benchmarks and insights from ptb-xl, 2020.

[15] Irem Sayin, Rana Gursoy, Buse Cicek, Yunus Emre Mert, Fatih Ozturk, Taha Emre Pamukcu, Ceylin Deniz Sevimli, and Huseyin Uvet. Cnn based detection of cardiovascular diseases from ecg images, 2024.

[16] Beyazit Bestami Yuksel and Ayse Yilmazer Metin. Advancing biomedical signal security: Real-time ecg monitoring with chaotic encryption, 2024.

[17] Atit Pokharel, Shashank Dahal, Pratik Sapkota, and Bhupendra Bimal Chhetri. Electrocardiogram (ecg) based cardiac arrhythmia detection and classification using machine learning algorithms, 2024.

[18] Tomás Teijeiro, Constantino A. García, Daniel Castro, and Paulo Félix. Arrhythmia classification from the abductive interpretation of short single-lead ecg records, 2017.

[19] Hao Tung, Chao Zheng, Xinsheng Mao, and Dahong Qian. Multi-lead ecg classification via an information-based attention convolutional neural network, 2020.

[20] Deepta Rajan and Jayaraman J. Thiagarajan. A generative modeling approach to limited channel ecg classification, 2018.

[21] Adrian Atienza, Jakob Bardram, and Sadasivan Puthusserypady. Subject-based non-contrastive self-supervised learning for ecg signal processing, 2023.

[22] Ismail Sadiq, Erick A. Perez-Alday, Amit J. Shah, Ali Bahrami Rad, Reza Sameni, and Gari D. Clifford. Mythological medical machine learning: Boosting the performance of a deep learning medical data classifier using realistic physiological models, 2021.

[23] Francesco Girlanda, Olga Demler, Bjoern Menze, and Neda Davoudi. Enhancing cardiovascular disease prediction through multi-modal self-supervised learning, 2024.

[24] Jielin Qiu, Jiacheng Zhu, Shiqi Liu, William Han, Jingqi Zhang, Chaojing Duan, Michael Rosenberg, Emerson Liu, Douglas Weber, and Ding Zhao. Automated cardiovascular record retrieval by multimodal learning between electrocardiogram and clinical report, 2023.

[25] Zhengyao Ding, Yujian Hu, Youyao Xu, Chengchen Zhao, Ziyu Li, Yiheng Mao, Haitao Li, Qian Li, Jing Wang, Yue Chen, Mengjia Chen, Longbo Wang, Xuesen Chu, Weichao Pan, Ziyi Liu, Fei Wu, Hongkun Zhang, Ting Chen, and Zhengxing Huang. Large-scale cross-modality pretrained model enhances cardiovascular state estimation and cardiomyopathy detection from electrocardiograms: An ai system development and multi-center validation study, 2024.

[26] Gabriel Ott, Yannik Schaubelt, Juan Miguel Lopez Alcaraz, Wilhelm Haverkamp, and Nils Strodthoff. Using explainable ai to investigate electrocardiogram changes during healthy aging – from expert features to raw signals, 2024.

[27] Giovanny Barbosa Casanova, Darwin Orlando Cardozo Sarmiento, Mario Joaquin Illera Bustos, Andrés Orozco Duque, and Henry Andrade Caicedo. Techniques of acquisition and processing of electrocardiographic signals in the detection of cardiac arrhythmias. *Respuestas*, 24(2):91–102, 2019.

[28] Sajad Mousavi, Fatemeh Afghah, Fatemeh Khadem, and U. Rajendra Acharya. Ecg language processing (elp): a new technique to analyze ecg signals, 2020.

[29] S. Karpagachelvi, M. Arthanari, and M. Sivakumar. Ecg feature extraction techniques - a survey approach, 2010.

[30] Kratika Tyagi and Prof. Sanjeev Thakur. A survey on various data mining techniques for ecg meta analysis, 2016.

[31] Patrick Wagner, Temesgen Mehari, Wilhelm Haverkamp, and Nils Strodthoff. Explaining deep learning for ecg analysis: Building blocks for auditing and knowledge discovery, 2024.

[32] Selcan Kaplan Berkaya, Alper Kursat Uysal, Efnan Sora Gunal, Semih Ergin, Serkan Gunal, and M Bilginer Gulmezoglu. A survey on ecg analysis. *Biomedical Signal Processing and Control*, 43:216–235, 2018.

[33] Hyunseung Chung, Jiho Kim, Joon myoung Kwon, Ki-Hyun Jeon, Min Sung Lee, and Edward Choi. Text-to-ecg: 12-lead electrocardiogram synthesis conditioned on clinical text reports, 2023.

[34] Che Liu, Zhongwei Wan, Sibo Cheng, Mi Zhang, and Rossella Arcucci. Etp: Learning transferable ecg representations via ecg-text pre-training, 2023.

[35] Amnon Bleich, Antje Linnemann, Bjoern H. Diem, and Tim OF Conrad. Automated medical report generation for ecg data: Bridging medical text and signal processing with deep learning, 2024.

[36] Matthew A. Reyna, Deepanshi, James Weigle, Zuzana Koscova, Kiersten Campbell, Kshama Kodthalu Shivashankara, Soheil Saghafi, Sepideh Nikookar, Mohsen Motie-Shirazi, Yashar Kiarashi, Salman Seyedi, Gari D. Clifford, and Reza Sameni. Ecg-image-database: A dataset of ecg images with real-world imaging and scanning artifacts; a foundation for computerized ecg image digitization and analysis, 2024.

[37] Divya Shanmugam, Davis Blalock, and John Guttag. Multiple instance learning for ecg risk stratification, 2020.

[38] Samrajya Thapa, Koushik Howlader, Subhankar Bhattacharjee, and Wei le. More: Multi-modal contrastive pre-training with transformers on x-rays, ecgs, and diagnostic report, 2024.

[39] Taminul Islam, Arindom Kundu, Tanzim Ahmed, and Nazmul Islam Khan. Analysis of arrhythmia classification on ecg dataset, 2023.

[40] B S Chandra, C S Sastry, and S Jana. Robust heartbeat detection from multimodal data via cnn-based generalizable information fusion, 2018.

[41] Eran Zvuloni, Jesse Read, Antônio H. Ribeiro, Antonio Luiz P. Ribeiro, and Joachim A. Behar. On merging feature engineering and deep learning for diagnosis, risk-prediction and age estimation based on the 12-lead ecg, 2022.

[42] Jintai Chen, Kuanlun Liao, Kun Wei, Haochao Ying, Danny Z. Chen, and Jian Wu. Me-gan: Learning panoptic electrocardio representations for multi-view ecg synthesis conditioned on heart diseases, 2023.

[43] Nhat-Tan Bui, Dinh-Hieu Hoang, Thinh Phan, Minh-Triet Tran, Brijesh Patel, Donald Adjeroh, and Ngan Le. Tsrnet: Simple framework for real-time ecg anomaly detection with multimodal time and spectrogram restoration network, 2024.

[44] Md Manjurul Ahsan and Zahed Siddique. Machine learning-based heart disease diagnosis: A systematic literature review, 2021.

[45] Shourya Verma. Development of interpretable machine learning models to detect arrhythmia based on ecg data, 2022.

[46] Zibin Zhao. Transforming ecg diagnosis:an in-depth review of transformer-based deeplearning models in cardiovascular disease detection, 2023.

[47] Han Yu, Peikun Guo, and Akane Sano. Ecg semantic integrator (esi): A foundation ecg model pretrained with llm-enhanced cardiological text, 2024.

[48] Weijie Sun, Sunil Vasu Kalmady, Amir Salimi, Nariman Sepehrvand, Eric Ly, Abram Hindle, Russell Greiner, and Padma Kaul. Ecg for high-throughput screening of multiple diseases: Proof-of-concept using multi-diagnosis deep learning from population-based datasets, 2022.

[49] Jialu Tang, Tong Xia, Yuan Lu, Cecilia Mascolo, and Aaqib Saeed. Electrocardiogram-language model for few-shot question answering with meta learning, 2024.

[50] R Karthik, Dhruv Tyagi, Amogh Raut, Soumya Saxena, and Rajesh Kumar M. Implementation of neural network and feature extraction to classify ecg signals, 2018.

[51] Temesgen Mehari and Nils Strodthoff. Self-supervised representation learning from 12-lead ecg data, 2022.

[52] Jialu Tang, Tong Xia, Yuan Lu, Cecilia Mascolo, and Aaqib Saeed. Electrocardiogram report generation and question answering via retrieval-augmented self-supervised modeling, 2024.

[53] Seokmin Choi, Sajad Mousavi, Phillip Si, Haben G. Yhdego, Fatemeh Khadem, and Fatemeh Afghah. Ecgbert: Understanding hidden language of ecgs with self-supervised representation learning, 2023.

[54] Jungwoo Oh, Hyunseung Chung, Joon myoung Kwon, Dong gyun Hong, and Edward Choi. Lead-agnostic self-supervised learning for local and global representations of electrocardiogram, 2022.

[55] Sahar Soltanieh, Ali Etemad, and Javad Hashemi. Analysis of augmentations for contrastive ecg representation learning, 2022.

[56] Jielin Qiu, William Han, Jiacheng Zhu, Mengdi Xu, Michael Rosenberg, Emerson Liu, Douglas Weber, and Ding Zhao. Transfer knowledge from natural language to electrocardiography: Can we detect cardiovascular disease through language models?, 2023.

[57] V. V. Kuznetsov, V. A. Moskalenko, and N. Yu. Zolotykh. Electrocardiogram generation and feature extraction using a variational autoencoder, 2020.

[58] Viktor van der Valk, Douwe Atsma, Roderick Scherptong, and Marius Staring. Joint optimization of a β-vae for ecg task-specific feature extraction, 2023.

[59] Christopher J. Harvey, Sumaiya Shomaji, Zijun Yao, and Amit Noheria. Comparison of autoencoder encodings for ecg representation in downstream prediction tasks, 2024.

[60] icardo: A machine learning based smart healthcare framework for cardiovascular disease prediction.

[61] Shenda Hong, Yuxi Zhou, Junyuan Shang, Cao Xiao, and Jimeng Sun. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review, 2020.

[62] Maximilian P Oppelt, Maximilian Riehl, Felix P Kemeth, and Jan Steffan. Combining scatter transform and deep neural networks for multilabel electrocardiogram signal classification, 2020.

[63] Ya Zhou, Xiaolin Diao, Yanni Huo, Yang Liu, Xiaohan Fan, and Wei Zhao. Masked transformer for electrocardiogram classification, 2024.

[64] Morteza Maleki and Foad Haeri. Identification of cardiovascular diseases through ecg classification using wavelet transformation, 2024.

[65] Li Xiaolin, Fang Xiang, Rajesh C. Panicker, Barry Cardiff, and Deepu John. Classification of ecg based on hybrid features using cnns for wearable applications, 2022.

[66] Ding Zhu, Vishnu Kabir Chhabra, and Mohammad Mahdi Khalili. Ecg signal denoising using multi-scale patch embedding and transformers, 2024.

[67] Zeeshan Ahmad, Anika Tabassum, Ling Guan, and Naimul Khan. Ecg heartbeat classification using multimodal fusion, 2021.

[68] Cheng Guo, Sajid Ahmed, and Mohamed-Slim Alouini. Machine learning-based automatic cardiovascular disease diagnosis using two ecg leads, 2023.

[69] Radhika Dua, Jiyoung Lee, Joon myoung Kwon, and Edward Choi. Automatic detection of noisy electrocardiogram signals without explicit noise labels, 2022.

[70] Dongdong Zhang, Xiaohui Yuan, and Ping Zhang. Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram, 2020.

[71] Yubao Zhao, Tian Zhang, Xu Wang, Puyu Han, Tong Chen, Linlin Huang, Youzhu Jin, and Jiaju Kang. Ecg-chat: A large ecg-language model for cardiac disease diagnosis, 2024.

[72] Marko Velic, Ivan Padavic, and Sinisa Car. Computer aided ecg analysis - state of the art and upcoming challenges, 2013.

[73] Nabil Ibtehaz and Masood Mortazavi. Modally reduced representation learning of multi-lead ecg signals through simultaneous alignment and reconstruction, 2024.

[74] Cristina Rueda, Yolanda Larriba, and Adrián Lamela. The hidden waves in the ecg uncovered: a sound automated interpretation method, 2020.

[75] Manh Pham, Aaqib Saeed, and Dong Ma. C-melt: Contrastive enhanced masked auto-encoders for ecg-language pre-training, 2024.

[76] Prapti Ganguly, Wazib Ansar, and Amlan Chakrabarti. Enhancing electrocardiogram signal analysis using nlp-inspired techniques: A novel approach with embedding and self-attention, 2024.

[77] Ramith Hettiarachchi, Udith Haputhanthri, Kithmini Herath, Hasindu Kariyawasam, Shehan Munasinghe, Kithmin Wickramasinghe, Duminda Samarasinghe, Anjula De Silva, and Chamira U. S. Edussooriya. A novel transfer learning-based approach for screening pre-existing heart diseases using synchronized ecg signals and heart sounds, 2021.

[78] Neha Soorma, Jaikaran Singh, and Mukesh Tiwari. Feature extraction of ecg signal using hht algorithm, 2014.

[79] Giacomo Lancia and Cristian Spitoni. Constructing interpretable prediction models with 1d dnns: An example in irregular ecg classification, 2024.

[80] Naoki Nonaka and Jun Seita. Data augmentation for electrocardiogram classification with deep neural network, 2020.

[81] Nils Strodthoff, Juan Miguel Lopez Alcaraz, and Wilhelm Haverkamp. Prospects for ai-enhanced ecg as a unified screening tool for cardiac and non-cardiac conditions – an explorative study in emergency care, 2024.

[82] Antônio H. Ribeiro, Manoel Horta Ribeiro, Gabriela Paixão, Derick Oliveira, Paulo R. Gomes, Jéssica A. Canazart, Milton Pifano, Wagner Meira Jr. au2, Thomas B. Schön, and Antonio Luiz Ribeiro. Automatic diagnosis of short-duration 12-lead ecg using a deep convolutional network, 2019.

[83] Simon Jaxy. Teaching a machine to diagnose a heart disease; beginning from digitizing scanned ecgs to detecting the brugada syndrome (brs), 2020.

[84] Khiem H. Le, Hieu H. Pham, Thao BT. Nguyen, Tu A. Nguyen, Tien N. Thanh, and Cuong D. Do. Lightx3ecg: A lightweight and explainable deep learning system for 3-lead electrocardiogram classification, 2022.

[85] Aristotelis Ballas and Christos Diou. A domain generalization approach for out-of-distribution 12-lead ecg classification with convolutional neural networks, 2022.

[86] Cuong V. Nguyen and Cuong D. Do. Transfer learning in ecg diagnosis: Is it effective?, 2024.

[87] Huy Pham, Konstantin Egorov, Alexey Kazakov, and Semen Budennyy. Machine learning-based detection of cardiovascular disease using ecg signals: performance vs. complexity, 2023.

[88] Tuija Leinonen, David Wong, Antti Vasankari, Ali Wahab, Ramesh Nadarajah, Matti Kaisti, and Antti Airola. Empirical investigation of multi-source cross-validation in clinical ecg classification, 2024.

[89] Péter Kovács, Carl Böck, Thomas Tschoellitsch, Mario Huemer, and Jens Meier. Diagnostic quality assessment for low-dimensional ecg representations, 2022.

[90] Ali Raza, Kim Phuc Tran, Ludovic Koehl, and Shujun Li. Designing ecg monitoring healthcare system with federated transfer learning and explainable ai, 2022.

[91] Felix Krones, Ben Walker, Terry Lyons, and Adam Mahdi. Combining hough transform and deep learning approaches to reconstruct ecg signals from printouts, 2024.

[92] Cheng Ding, Tianliang Yao, Chenwei Wu, and Jianyuan Ni. Deep learning for personalized electrocardiogram diagnosis: A review, 2024.

[93] Temesgen Mehari, Ashish Sundar, Alen Bosnjakovic, Peter Harris, Steven E. Williams, Axel Loewe, Olaf Doessel, Claudia Nagel, Nils Strodthoff, and Philip J. Aston. Ecg feature importance rankings: Cardiologists vs. algorithms, 2023.

[94] Daniel Mauricio Jimenez Gutierrez, Hafiz Muuhammad Hassan, Lorella Landi, Andrea Vitaletti, and Ioannis Chatzigiannakis. Application of federated learning techniques for arrhythmia classification using 12-lead ecg signals, 2024.

[95] Linhai Ma and Liang Liang. Improve robustness of dnn for ecg signal classification:a noise-to-signal ratio perspective, 2021.

[96] Sergio González, Abel Ko-Chun Yi, Wan-Ting Hsieh, Wei-Chao Chen, Chun-Li Wang, Victor Chien-Chia Wu, and Shang-Hung Chang. Multi-modal heart failure risk estimation based on short ecg and sampled long-term hrv, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.