Query Details

Back to Main Page

1. Kindly check and verify that the author name(s) and the identification of the corresponding author(s) are correctly recognized and presented in the correct sequence order and spellings, i.e., given name, middle name/initial, and family name for the authors. In addition, please verify that the E-mail addresses and Affiliation(s) of the corresponding author(s) and co-author(s) shown on the metadata page are valid, and make any necessary amendments if required.

2. Figure text size was too small in Figs. 1–14. Please provide high-resolution source file for these figures.

Channel Efficiency in ECG-Based Multi-label Cardiac Diagnosis

Dmitry Alexandrov
Email : dmytro.bolshak@temple.edu
AQ1 Affiliationids : Aff1

Claudia Dumitrescu
Email : claudia-anca.dumitrescu@temple.edu
Affiliationids : Aff1

Sadia Purba
Email : sadia.afrin.purba@temple.edu
Affiliationids : Aff1

Anna Chau
Email : anna.chau@temple.edu
Affiliationids : Aff1

Joseph Picone ✉
Email : picone@temple.edu
Affiliationids : Aff1, Correspondingaffiliationid : Aff1

Aff1 The Neural Engineering Data Consortium, Temple University, Philadelphia, PA, USA

Abstract

Electrocardiogram (ECG) analysis relies on preprocessed data and derived channels, which are linear combinations of raw signals, to aid in cardiac diagnosis. Medical professionals have developed these methods over the years, incorporating both clinical experience and technical knowledge of their monitoring equipment. The widespread use of these preprocessing steps has become standard practice in cardiology. However, recent advances in deep learning necessitate a reevaluation of these preprocessing steps. Deep learning models have demonstrated the ability to analyze raw signal or image data without requiring traditional preprocessing or derived channels. For example, architectures such as ResNet18 have achieved high performance on ECG datasets such as the Telehealth Network of Minas Gerais (TNMG) CODE Corpus. These architectures challenge the clinical assumption that derived channels are essential for accurate diagnosis. This shift raises the question of whether preprocessing methods developed for visual interpretation by clinicians remain necessary in machine learning contexts. In this chapter, we investigate a variety of baseline models for multi-label ECG classification. Our findings reveal that the performance of complex architectures, such as ResNet18, consistently declines as the number of input channels increases. We demonstrate that the original 8 channels collected using a standard 10-lead ECG can provide optimal performance on the TNMG Corpus. We describe the experimental setup used to compare models trained on raw ECG signals versus those trained on derived or preprocessed inputs. The results demonstrate that using raw signals not only simplifies the data pipeline during training but also preserves the native morphology of waveforms and outperforms diagnostic models relying on engineered features. Overall, learning directly from raw ECG data streamlines model development, reduces computational overhead, improves robustness across varying conditions, and enhances interpretability by eliminating opaque preprocessing stages.

Keywords

Cardiac diagnosis
ECG
Multi-label classification
Channel reduction
Deep learning

1. Introduction

The electrocardiogram (ECG) machine was invented by Willem Einthoven in 1903 [1]. For this groundbreaking work, Einthoven was awarded the Nobel Prize in Physiology or Medicine in 1924. Automated clinical ECG interpretation emerged in the early 1970s, marked by Ray Bonner's development of the IBM 5880 system, which utilized third-generation DSP boards to enable real time, on-site QRS complex detection [2]. Major device manufacturers quickly followed suit, and by the mid-1970s [3], manufacturers such as Philips and IBM had integrated proprietary QRS detection firmware into bedside monitors. These systems featured dual-filter pipelines to improve complex signal classification across multiple leads. In the early 1990s, the adoption of discrete wavelet transform (DWT) analysis [4] gained traction, offering multiresolution feature extraction that significantly enhanced performance—raising QRS detection accuracy above 99.8% on the MIT/BIH database. By the 2020s, deep learning (DL) approaches using big data resources had become commonplace in many medical applications, suggesting that it makes sense to revisit how ECG signals are preprocessed [5, 6].

The progression of ECG technology from Holter's early prototypes [7] to modern high-resolution recorders reflects a long-standing effort to increase data richness through added leads and higher sampling rates. By the 1960s, commercial multi-lead Holter systems recorded 3 to 12 leads continuously. While these improvements aimed to preserve signal fidelity, studies have shown that only a limited number of principal components are needed to capture most of the diagnostic information in multi-lead ECGs [8], while the addition of other leads adds noise. This implies that beyond a certain threshold, additional channels and higher sample rates contribute little new information, instead inflating storage and processing costs without proportional diagnostic gain.

In fact, today, there is great interest in wearable technology that does continuous monitoring from a single-lead [9] and low-cost systems that use a reduced number of leads [10]. Though the focus of this work is not an exploration of minimizing the number of leads for low-cost consumer devices or simplification of clinical equipment, our findings obviously have clinical and consumer implications. However, our focus here is more about how deep learning systems now offer the potential to simplify the interpretation of standard clinical ECG systems.

Access to state-of-the-art cardiovascular healthcare remains a critical issue in developing countries, primarily due to the severe shortage of cardiologists. For instance, in Brazil in 2022, there were only 8.40 cardiologists for every 100,000 individuals [11], while in Africa, a continent with a population of approximately 1.2 billion people, there are only about 2,000 recorded cardiologists, or 0.17 cardiologists per 100,000 individuals [12]. This alarming scarcity of official cardiologists requires the urgent need for innovative solutions, and although low-cost ECG machines are available in these regions, they fall short in providing comprehensive diagnoses due to their reliance on classical and naïve algorithms. These machines often lack the capacity to comprehend a wider range of present cardiovascular disease indicators, reducing their effectiveness in detecting indicators of cardiovascular diseases.

Hence, there is great interest in developing low-cost automated technology that can diagnose a broad range of conditions. Fortunately, DL technology has progressed to the point where this is now feasible provided adequate data exists to train such models. While there are several large publicly available datasets for medical applications [13, 14, 15], Ribeiro et al. [16, 17] have released a very significant corpus of ECGs that makes it possible to explore many dimensions of this problem space. This unique resource, which is well annotated and curated, is the focus of our work. In this chapter, we investigate the problem of signal preprocessing on the TNMG Corpus. However, there are many other interesting problems yet to be explored, such as the localization of critical events in the signal, often referred to as the micro-segmentation problem.

This chapter begins with a basic review of the signal processing steps used in ECG analysis in Sect. 2. We then discuss the problem of visual interpretation of an ECG signal in Sect. 3. Next, we review a variety of approaches to detect anomalies in ECGs based on DL principles in Sect. 4. In Sect. 5, we discuss the TNMG Corpus that we have used in our studies and review some of the work done on this impressive database. In Sect. 6, we present the results of our channel reduction experiments. Finally, in Sect. 7, we present evidence demonstrating that substantial reductions in channel dimensionality can be achieved without compromising diagnostic fidelity, thereby challenging the continued relevance of traditional ECG processing pipelines. We conclude with some comments on future research directions.

2. Signal Processing in ECG Analysis

Standard clinical ECG recordings [18] employ ten electrodes, as shown in Fig. 1 AQ2, to collect eight signal channels. The postprocessing steps for these channels are shown in Fig. 2. Six precordial leads (V1–V6) are positioned on the chest, while four limb leads attach to the right arm (RA), left arm (LA), left leg (LL), and right leg (RL). The RL lead serves as ground. This configuration records eight raw waveforms: six from the precordial leads, plus DI (LA-RA potential difference) and DII (LL-RA potential difference).

Fig. 1

The physical locations of the 10 electrodes used to record a 12-lead ECG

Alt Text

Diagram illustrating anatomical positions and muscle activation during arm movements. The left side shows a human torso with labeled muscle groups: pectoralis major, latissimus dorsi, anterior deltoid, and posterior deltoid. The right side displays

six subfigures of a person performing arm movements with corresponding muscle activation graphs. Each subfigure highlights different angles and muscle engagement, with graphs indicating the level of activation for each muscle group.

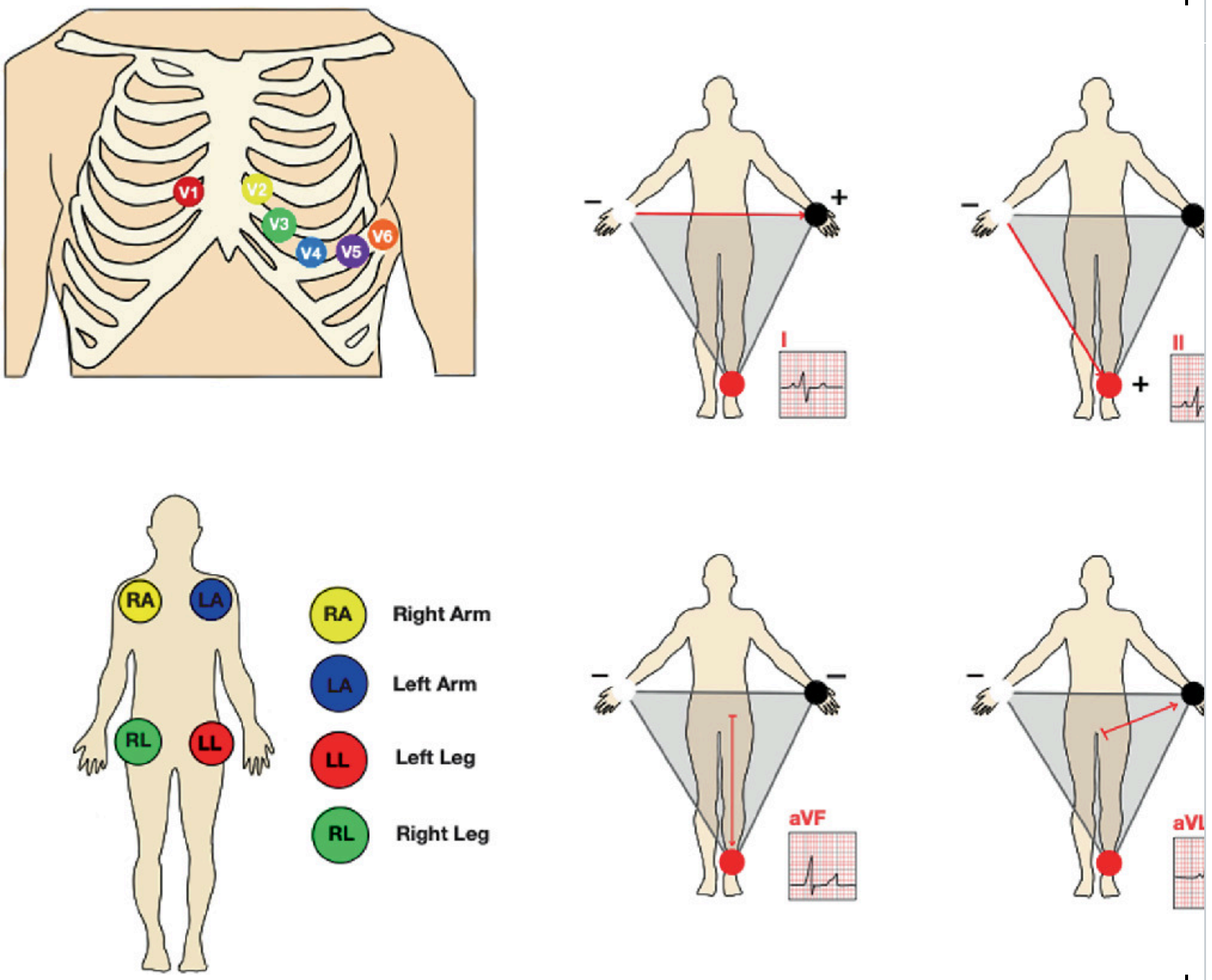
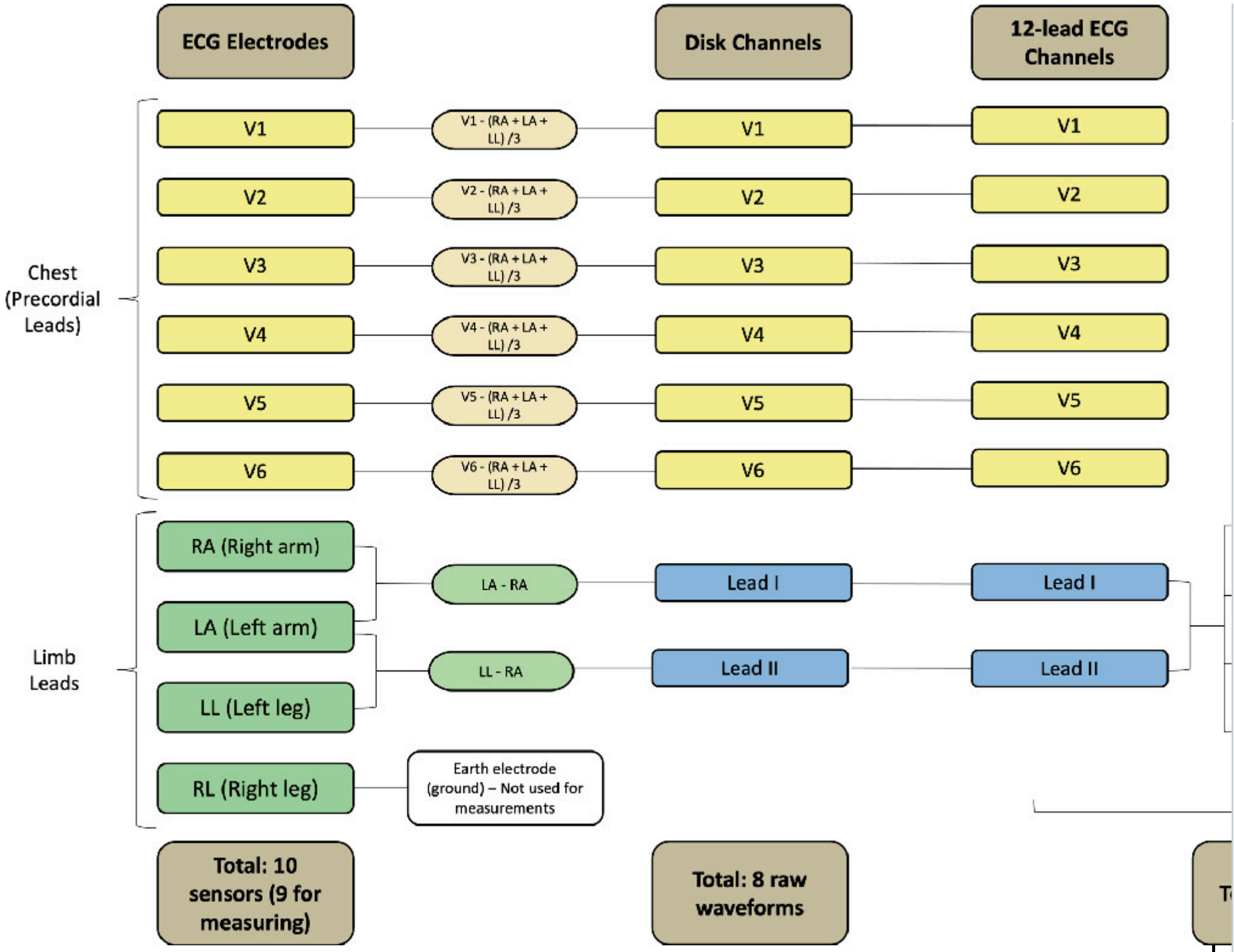


Fig. 2

Conversion of an ECG collected with 10 leads to 8 and 12-channel signals

Alt Text

Flow chart illustrating the setup of ECG electrodes and channels. The chart is divided into sections: ECG Electrodes, Disk Channels, and 12-lead ECG Channels. Chest (precordial) leads V1 to V6 are shown, with connections to corresponding disk and ECG channels. Limb leads include RA (right arm), LA (left arm), LL (left leg), and RL (right leg), with connections forming Lead I, Lead II, and Lead III. Augmented leads are derived from these, labeled as avR, avL, and avF. The chart notes a total of 10 sensors, 8 raw waveforms, and 12 leads.



Four additional leads are derived mathematically using simple differences and averages:

$$\mathrm{DIII} = \mathrm{DII} - \mathrm{DI}, \tag{1}$$

$$\mathrm{aVR} = \frac{(\mathrm{DI} + \mathrm{DII})}{2}, \tag{2}$$

$$\mathrm{aVL} = \frac{(\mathrm{DI} - \mathrm{DII})}{2}, \tag{3}$$

$$\mathrm{aVF} = \frac{(\mathrm{DII} - \mathrm{DI})}{2}. \tag{4}$$

These computations are summarized in Fig. 2 and have been part of the art of interpreting an ECG for many decades. In this chapter, we explore whether these additional channels are needed for modern DL systems.

The derived ECG leads are based on established electrophysiological principles that map the heart's electrical activity onto anatomical planes using potential differences measured at specific body locations. Leads DI, DII, and DIII form the basis of Einthoven's triangle [19], a conceptual model that approximates the frontal plane of the body as an equilateral triangle with vertices at the left arm (LA), right arm (RA), and left leg (LL). In this configuration, DI measures the potential difference between LA and RA, DII measures LL to RA, and DIII derives from the vector sum relationship in Eq. (1) reflecting Kirchhoff's voltage law around the triangle. This is depicted in Fig. 1.

Goldberger's augmented leads—aVR, aVL, and aVF—extend Einthoven's model by creating unipolar measurements with reference to a virtual ground. These leads synthesize additional frontal plane perspectives: aVR views the heart from the right shoulder, aVL from the left shoulder, and aVF from the foot. Equations (2)–(4) are derived from averaging limb lead signals to create reference points, enabling the visualization of cardiac depolarization along across multiple axes without adding physical electrodes.

These derived leads provide clinicians with complementary spatial views of cardiac activity, improving diagnostic resolution. However, they are linear transformations of the original limb leads and introduce redundancy in the signal set. In DL applications, where models can extract latent spatial and temporal patterns automatically, the inclusion of derived leads may be unnecessary or counterproductive. The anatomical rationale for their construction remains valid in clinical interpretation, but their utility in data-driven models must be empirically justified, particularly in the context of input redundancy, multicollinearity, and computational efficiency.

3. Visual Interpretation of Cardiology Signals

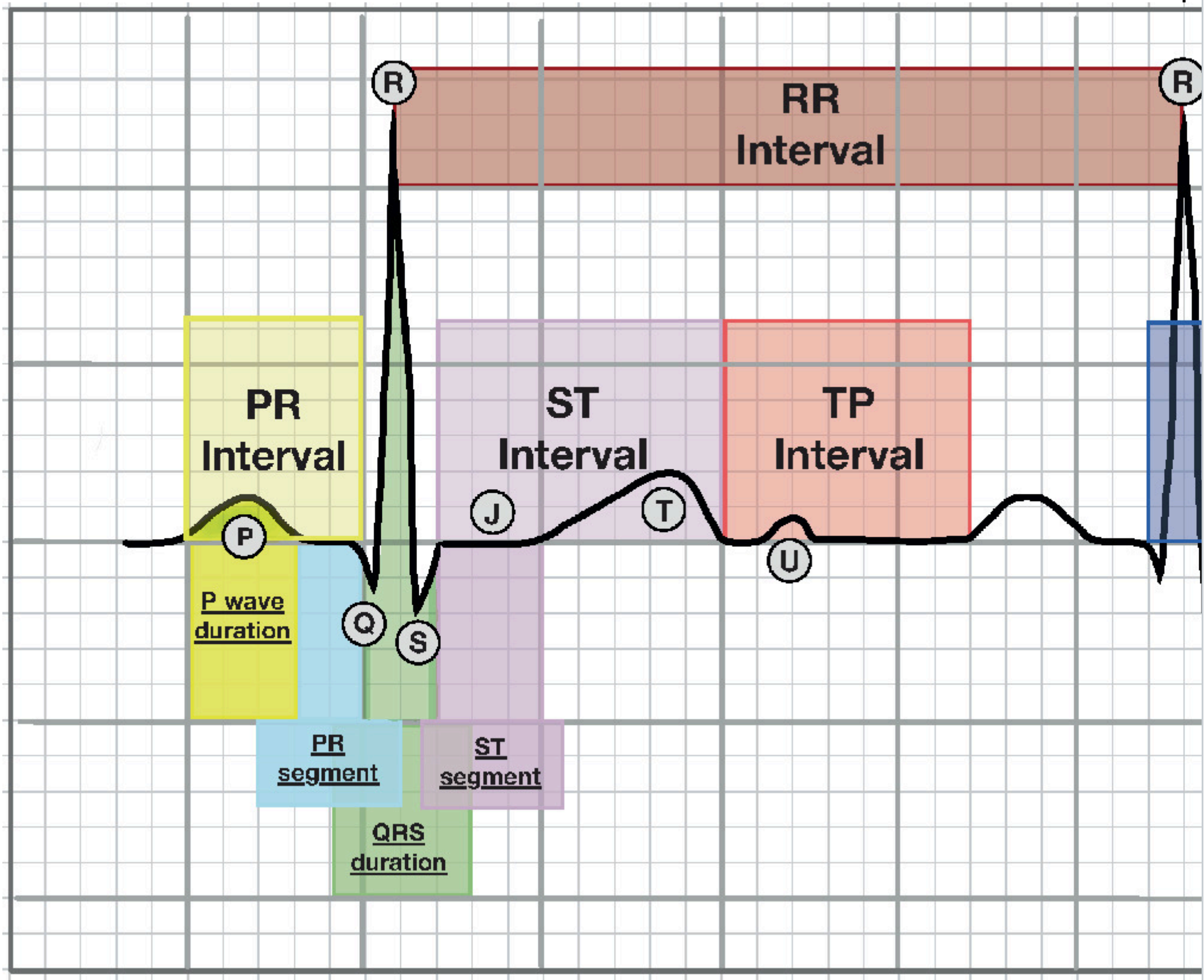
An electrocardiogram (ECG) captures the electrical activity of the heart via surface electrodes, providing a multichannel signal used for clinical diagnoses. A classical ECG curve is shown in Fig. 3. This is an illustration of a typical normal sinus rhythm [20] in which critical points on the waveform are labeled as P, Q, R, S, J, T, and U [18]. Each normal cardiac cycle produces a characteristic waveform consisting of a P wave (atrial depolarization), a QRS complex (ventricular depolarization), and a T-wave (ventricular repolarization). In a normal sinus rhythm, impulses originate in the sinoatrial node (a small cluster of specialized cardiac muscle cells located in the right atrium, near the opening of the superior vena cava), producing a regular narrow-complex ECG at roughly 60–100 beats per minute. In this pattern, each QRS complex is preceded by a single upright P wave, and the PR interval (onset of the P wave to onset of the QRS) remains constant. This configuration reflects optimal atrioventricular conduction (refers to the way electrical impulses travel from the atria to the ventricles of the heart through the atrioventricular node and related conduction pathways) and is often considered the benchmark for normal cardiac electrical activity. However, such pristine patterns are relatively uncommon in practice.

Fig. 3

A standard normal sinus rhythm with labeled P, QRS, and T waves

Alt Text

Electrocardiogram (ECG) waveform diagram illustrating various intervals and segments. The waveform includes labeled sections: PR Interval, QRS duration, ST Interval, QT Interval, and RR Interval. Each section is color-coded and marked with corresponding points such as P, Q, R, S, T, and U. The diagram provides time measurements for intervals and segments, with a speed of 25 mm/sec indicated.



Even among healthy individuals, deviations from this ideal are frequent. For instance, in a study [21] involving 26 healthy elderly men, 76.9% exhibited atrial ectopic beats (extra heartbeats that originate in the atria), and 76.9% had ventricular ectopic beats (extra heartbeats that originate in the ventricles instead of the sinoatrial node). Episodes of nonsustained ventricular tachycardia (abnormal heart rhythm that arises from the ventricles) were present in 11.5% of participants [21]. These findings suggest that transient

arrhythmias (abnormal heart rhythms that occur for a short period of time) are common even in populations without known cardiac disease.

In a standard 12-lead ECG [22], shown in Fig. 2, different ECG leads offer complementary spatial views. For example, septal depolarization (the first phase of ventricular depolarization in which an electrical impulse spreads through the interventricular septum—the wall separating the right and left ventricles) produces small Q waves in lateral leads (I, aVL, V5–V6) and the initial R in lead V1, whereas ventricular depolarization toward or away from each lead causes the characteristic positive or negative deflections. These normal features provide the baseline for identifying abnormalities.

A recent ML study by Ribeiro et al. [16, 17] targeted exactly six common ECG abnormalities using DL: (1) first-degree AV block (1dAVb), (2) right bundle branch block (RBBB), (3) left bundle branch block (LBBB), (4) sinus bradycardia, (5) atrial fibrillation, and (6) sinus tachycardia. The following paragraphs briefly describe each condition's definition, physiology, ECG manifestations (with primary leads), and clinical significance. A more extended description can be found in [17].

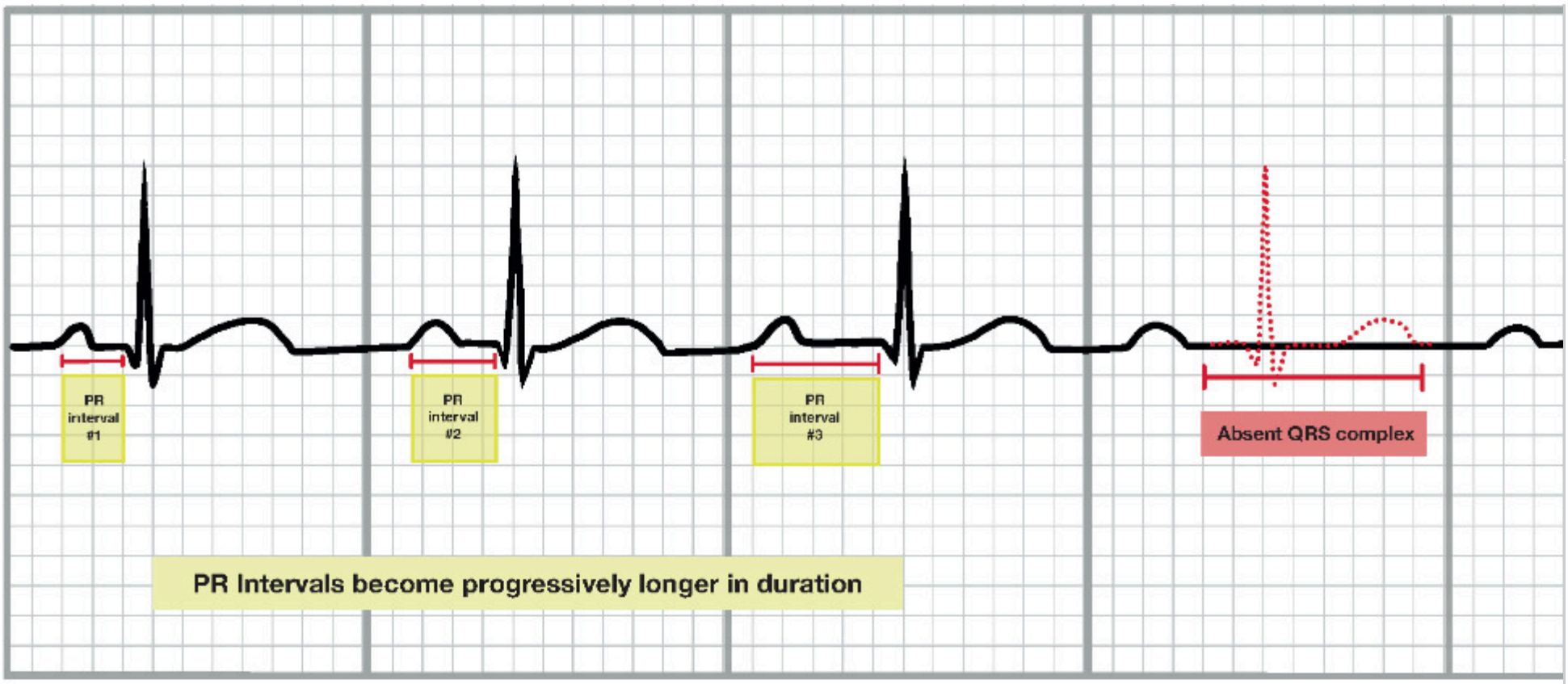
(1) *First-degree AV block (1dAVb)* [23], described in Fig. 4, is defined by uniformly prolonged conduction through the AV node. All atrial i lengthened delay. On the ECG, this appears as a PR interval exceeding $200ms$ while still maintaining one P wave before each QRS com (occasionally $> 300ms$) and is most readily measured in lead II. In terms of physiology, it often reflects increased AV nodal delay due t conduction system. It is frequently a benign finding seen in healthy young or athletic individuals and rarely causes symptoms [24]. Onl disease or medications (e.g., beta-blockers, digoxin, or Lyme carditis) does it warrant further evaluation.

Fig. 4

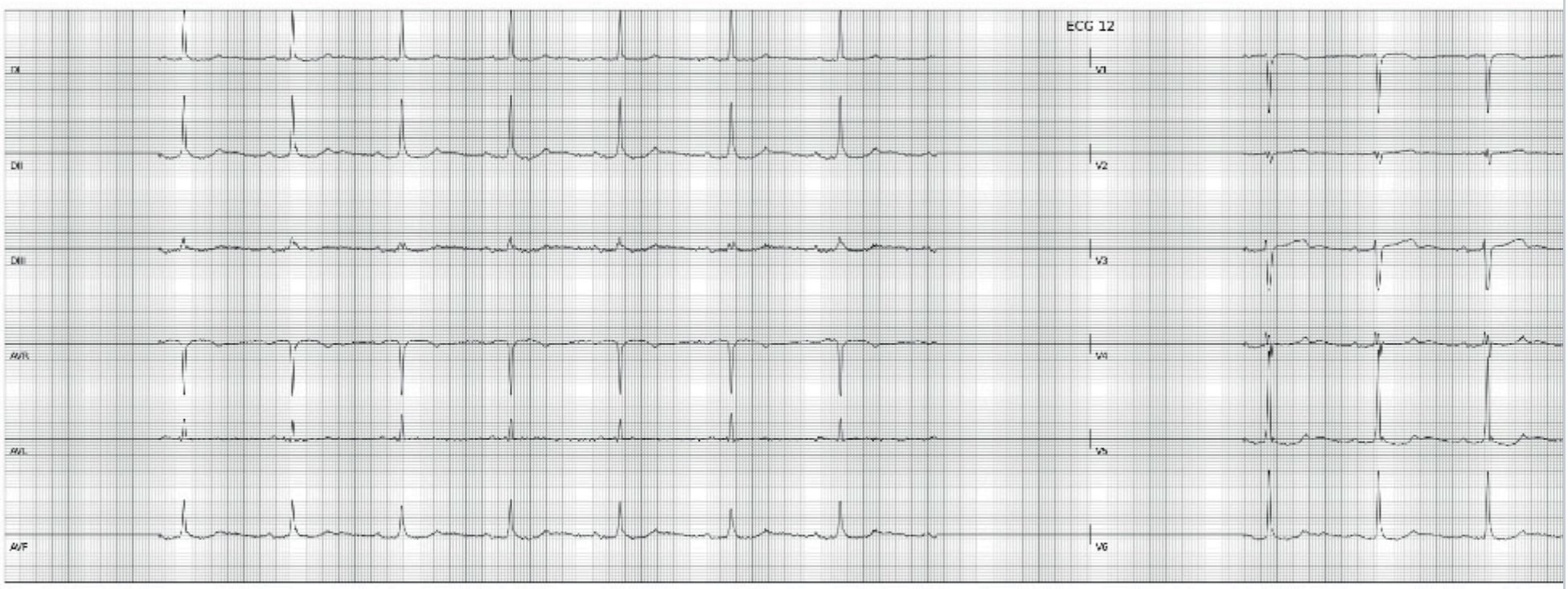
First-degree AV block (1adAVb) showing a prolonged PR interval

Alt Text

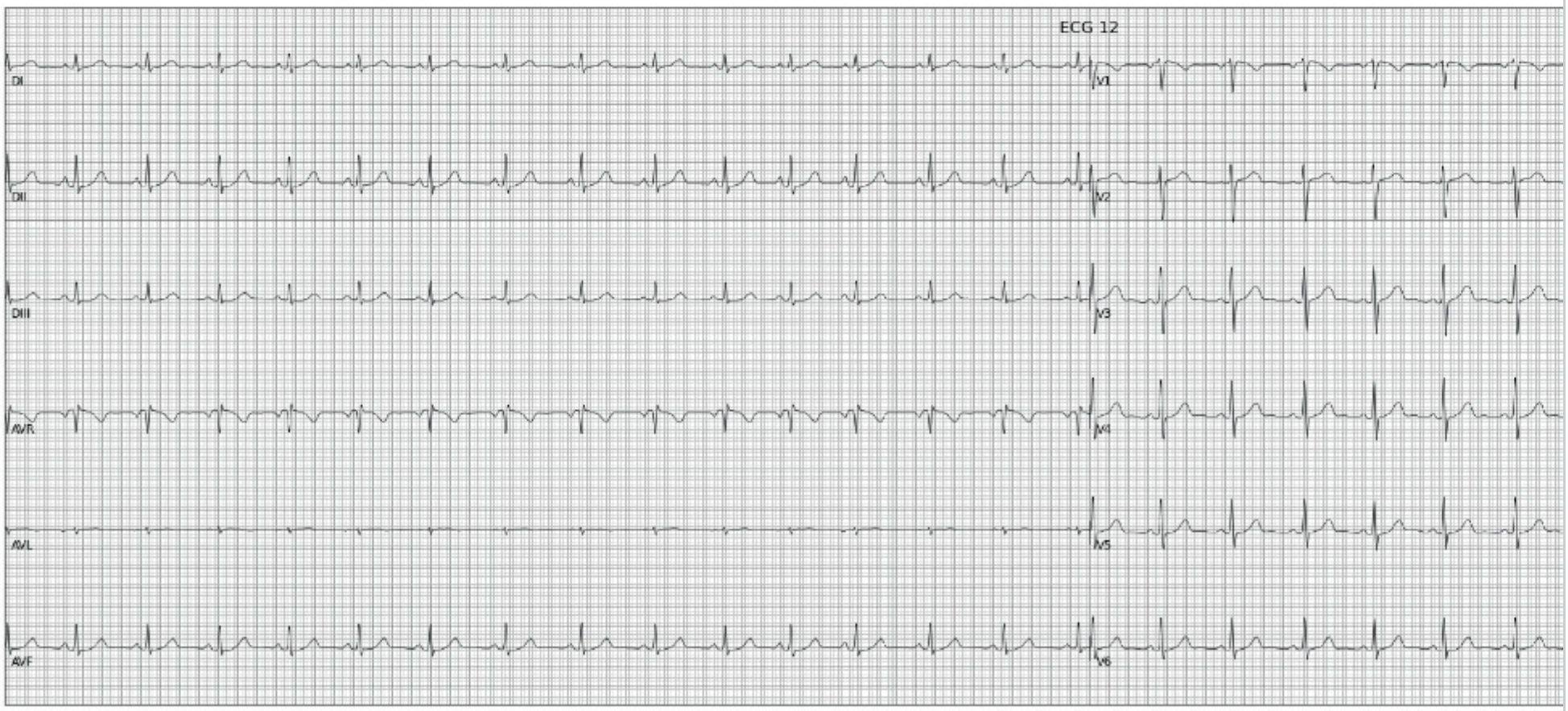
The image is a panel figure composed of three subfigures displaying electrocardiogram (ECG) signals. 1. The first subfigure (a) sh 1st degree AV block abnormality. Key features are highlighted with colored boxes and text. 2. The second subfigure (b) presents a though specific details are not highlighted. 3. The third subfigure (c) displays an ECG signal with no abnormalities present. Each s readings, showing the electrical activity of the heart over time.



(a) description of a 1dAVb abnormality



(b) abnormality present



(c) no abnormality present

(2) *Right bundle branch block (RBBB)* [25], described in Fig. 5. Right bundle branch block (RBBB) with rsR′ pattern in lead V1 and wide S conduction is delayed or interrupted. The left ventricle depolarizes normally, but the right ventricle must depolarize secondarily via slov

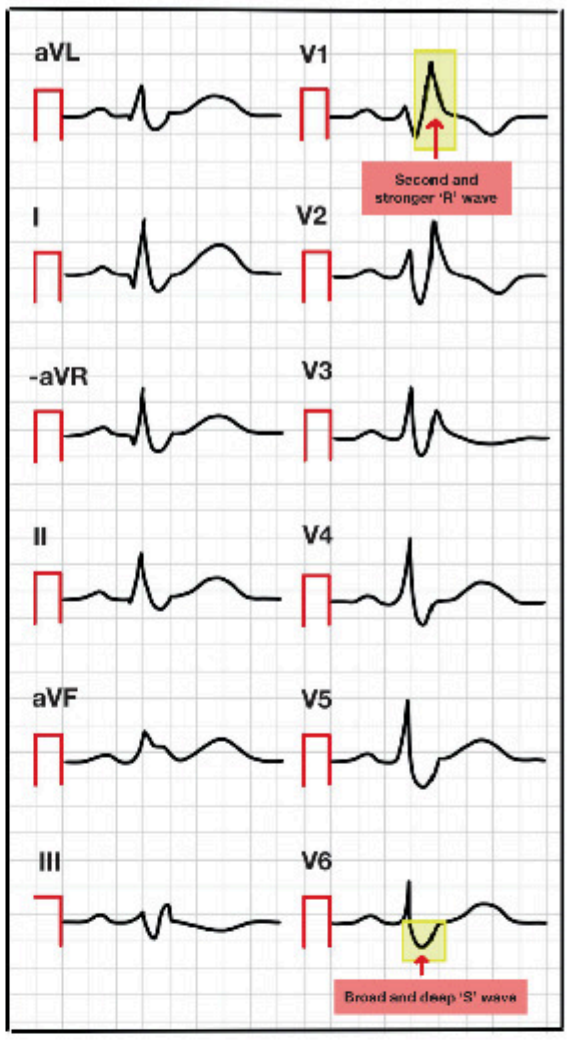
QRS (duration $> 120ms$) and a characteristic pattern on the ECG. In leads V1–V3, one sees an rsR′ (“M-shaped”) complex: a small ini right ventricular depolarization. In lateral leads (I, aVL, V5–V6), a broad, slurred S wave appears. For example, lead V1 shows an M-shap wide terminal S wave [26]. These changes reflect the normal septal (LV) activation followed by delayed RV activation. The cardiac axis Clinically, RBBBs may be a benign finding (even in athletes) or indicate right ventricular strain (e.g., pulmonary pathology) or underlying RBBB focus on the prolonged QRS and the distinctive V1–V6 morphologies. In practice, lead V1 (and V2) is examined for the “M” patter

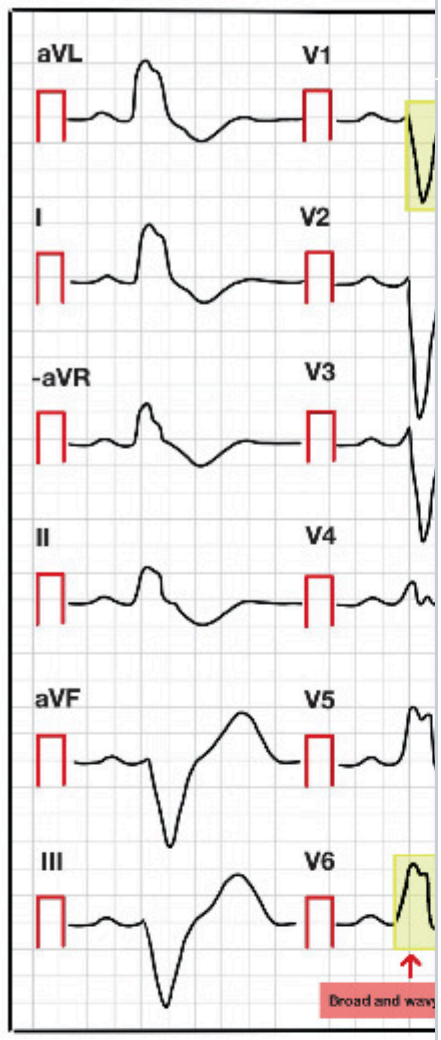
Fig. 5

Right bundle branch block (RBBB) with rsR′ pattern in lead V1 and wide S in V6

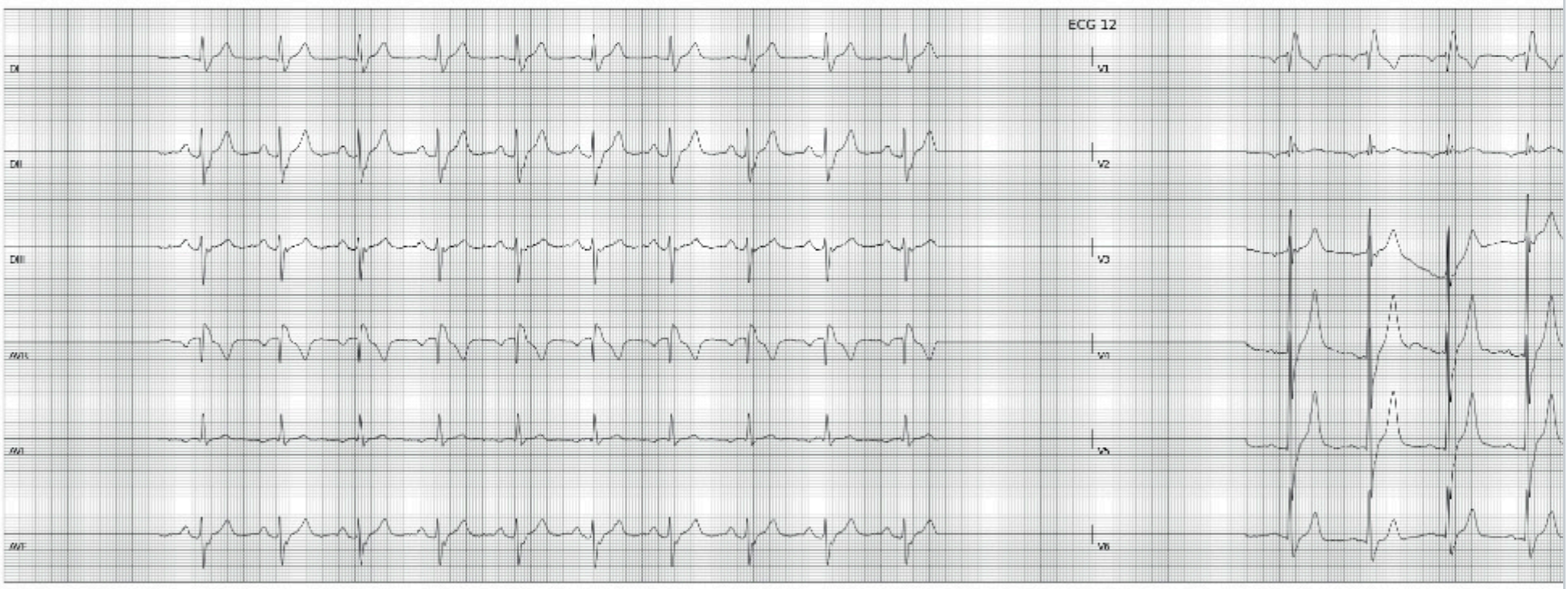
Alt Text

The image consists of three panels related to electrocardiogram (ECG) signals. Panel (a) provides a description of a right bundle specific waveform patterns. Panel (b) shows an ECG with the presence of an abnormality, likely RBBB, indicated by distinct wave abnormalities, showing regular waveform patterns. The panels are arranged to compare and contrast the different ECG signal cha

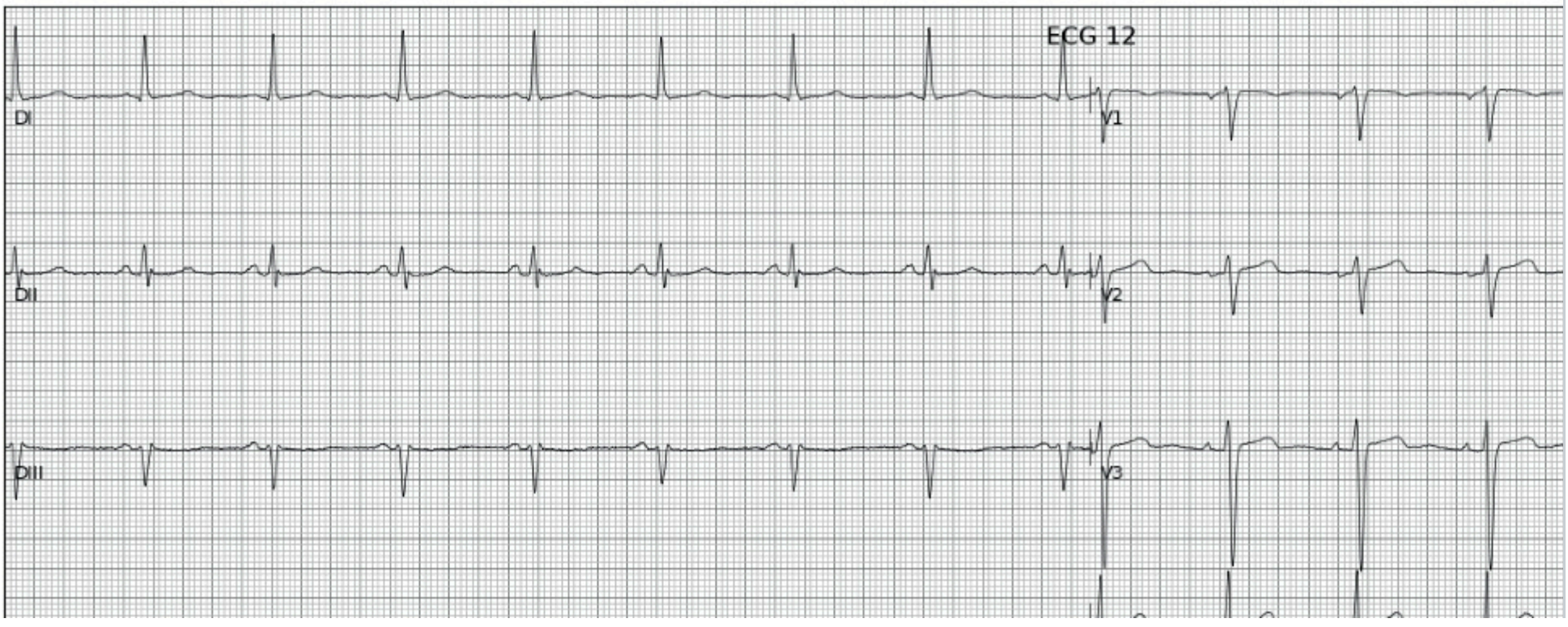




(a) description of a RBBB abnormality



(b) abnormality present



(c) no abnormality present

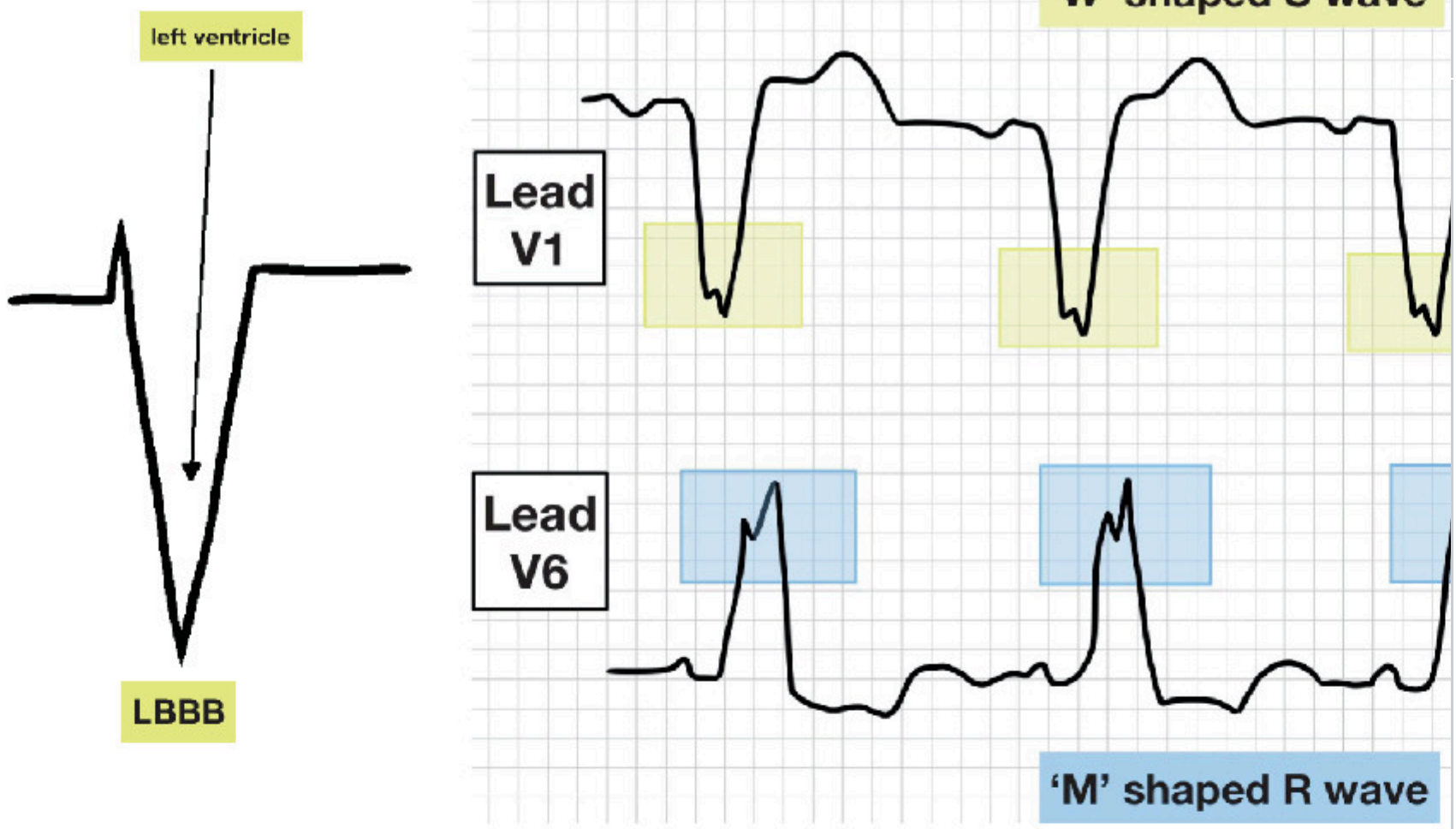
(3) *Left bundle branch block (LBBB)* [27], described in Fig. 6, involves delayed conduction in the left bundle. The impulse travels down th then spreads slowly to the left ventricle. This produces a widened QRS (>120 ms) with a markedly different morphology than RBBB [28 shaped) R-wave in lateral leads I, aVL, V5–V6, and a deep, wide S wave in right precordial leads V1–V3. The normal initial septal Q wave the overall QRS axis may shift leftward. Additional criteria are absent Q waves in V5–V6 and a prolonged R-wave peak time (>60 ms) in left ventricular depolarization is delayed. The result is tall R waves in V5–V6 and broad S waves in V1–V3. LBBB often signifies underlyi and complicates interpretation. An ML model must thus recognize LBBB as a distinct class of wide-QRS morphology. Clinically, leads V notched R) are inspected to confirm LBBB. A comparison of RBBB and LBBB is given in Fig. 5a.

Fig. 6

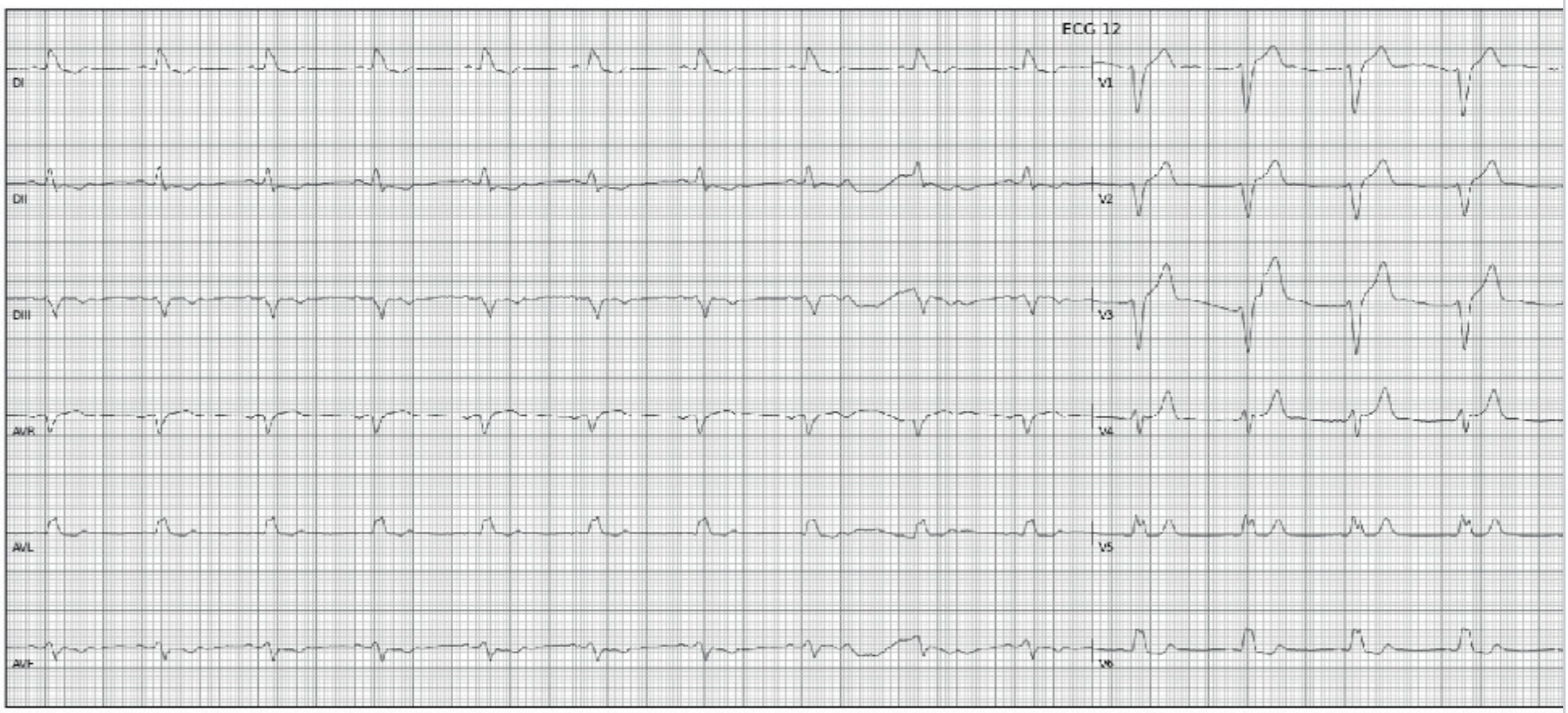
Left bundle branch block (LBBB) with broad notched R in V6 and deep S in V1

Alt Text

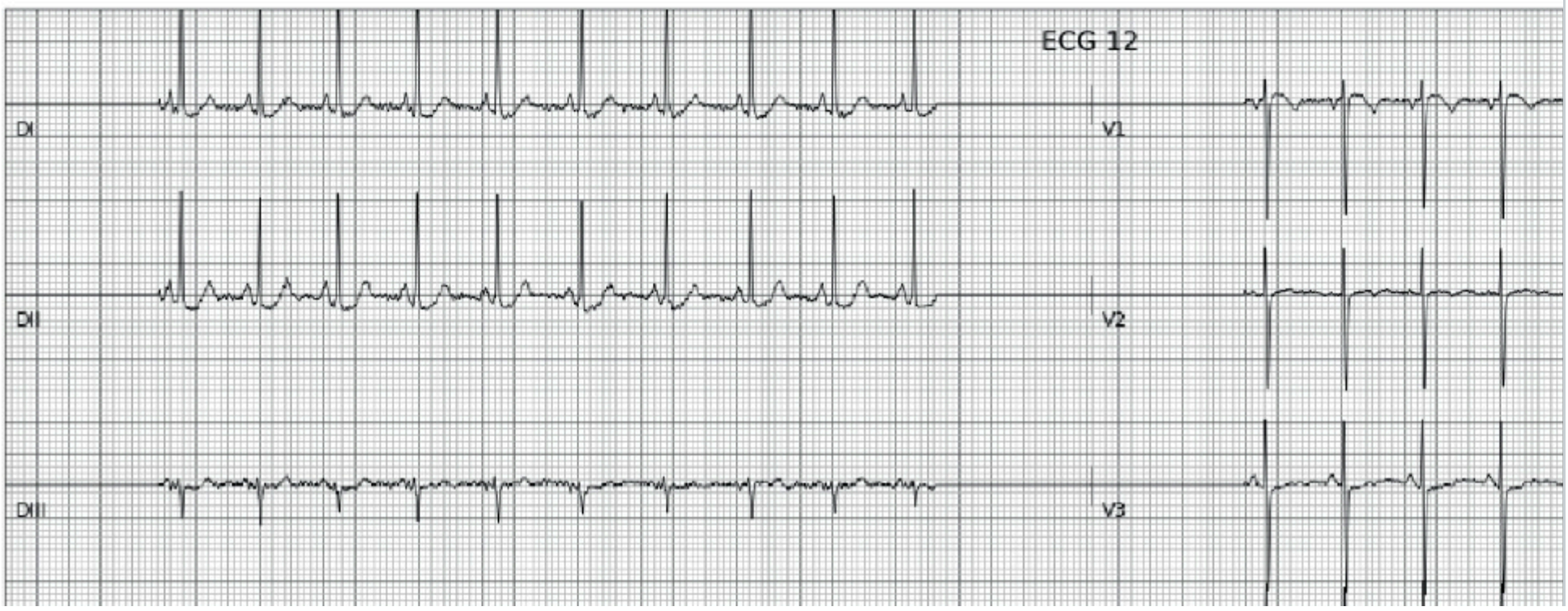
Diagram showing ECG signal patterns. The top section illustrates a normal QRS complex with labeled components, including a W wave in Lead V6. Below, three ECG strips are displayed: (a) shows a description of a left bundle branch block (LBBB) abnormality and (c) shows no abnormality. The image highlights differences in waveforms and patterns associated with cardiac conduction.



(a) description of a LBBB abnormality



(b) abnormality present



(c) no abnormality present

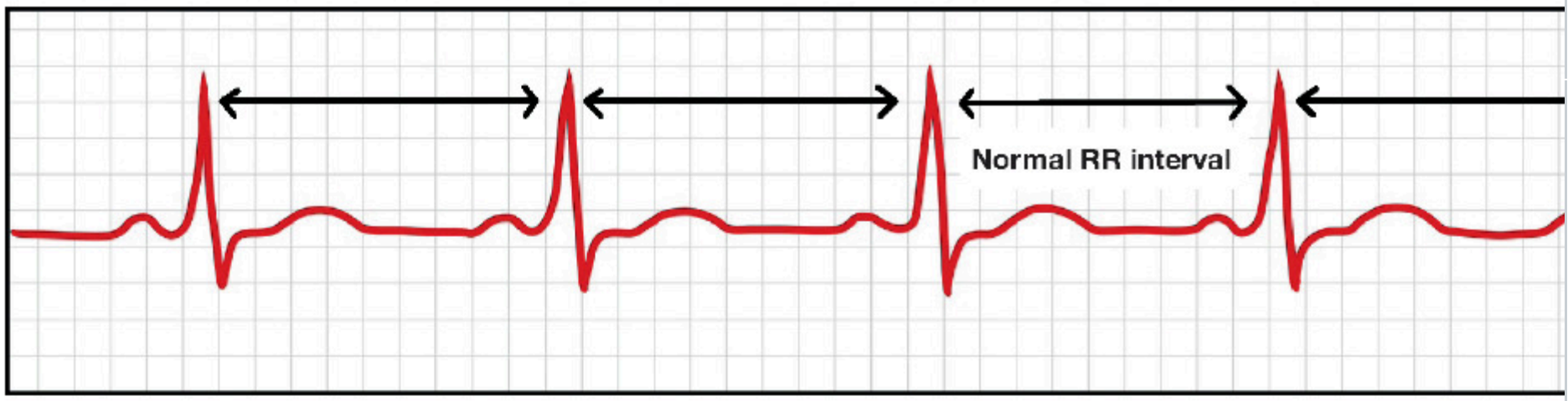
(4) *Sinus bradycardia (SB)* [29], described in Fig. 7, is simply a sinus rhythm with a slow rate. It is defined as a resting ventricular rate be the ECG features of sinus rhythm remain intact: there is a P wave before every QRS, the P wave morphology is normal (upright in leads difference is the longer R–R interval reflecting the slower rate. Often, prominent U waves (small deflections following the T-wave) may bradycardia. Sinus bradycardia commonly occurs physiologically during sleep or in well-trained athletes (reflecting high vagal tone) an from medications (beta-blockers, calcium-channel blockers) or conditions like hypothyroidism. In terms of diagnosis, lead II usually dis rhythm ($< 60bpm$) with one P-QRS-T sequence per cycle. ML systems classify sinus bradycardia by recognizing the slow but otherwi

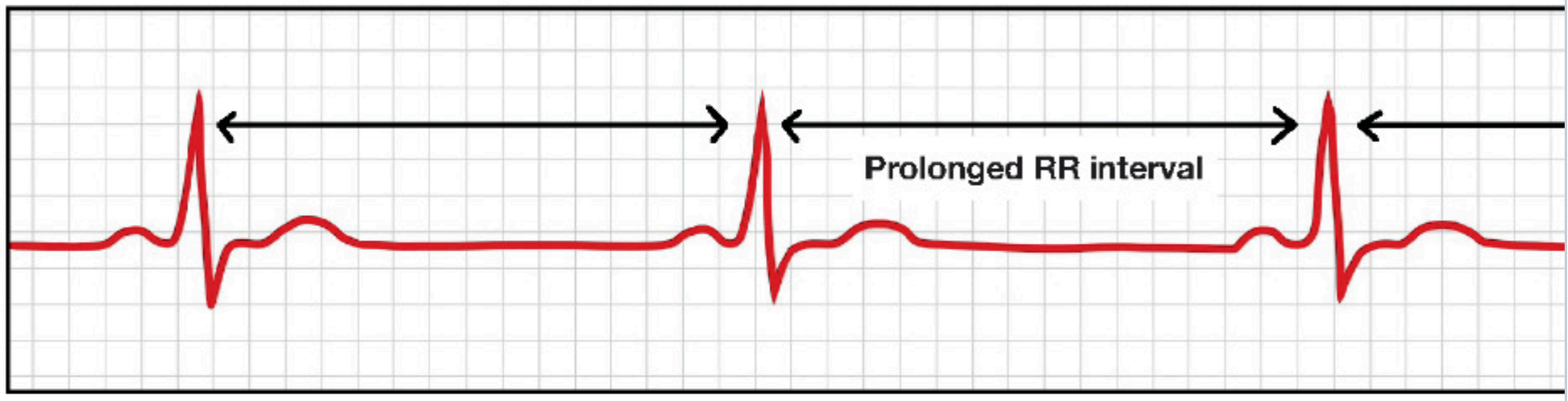
Fig. 7

Sinus bradycardia (SB) in which a normal upright P wave precedes every QRS complex

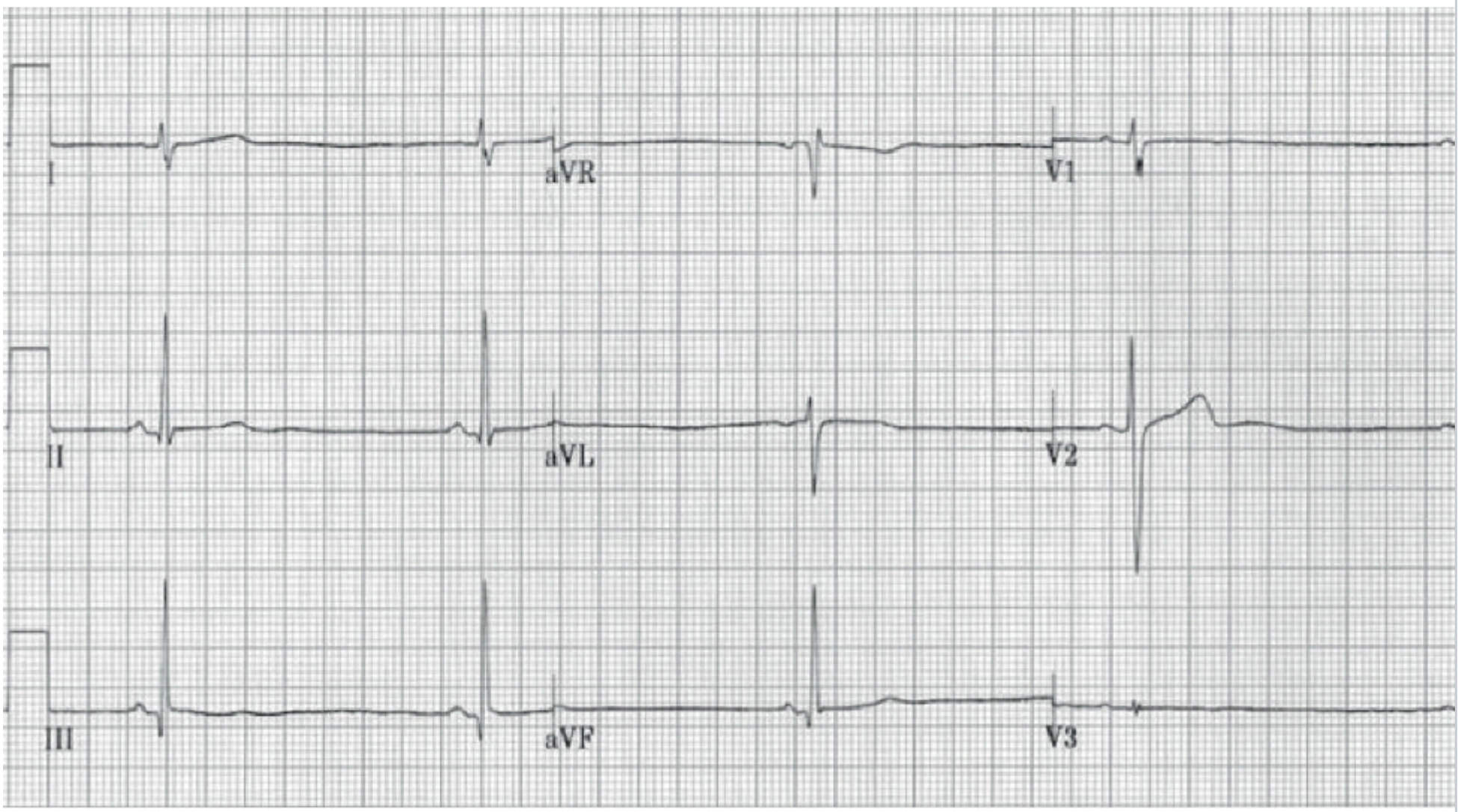
Alt Text

The image consists of two panels comparing electrocardiogram (ECG) signals. The top panel shows a normal sinus rhythm with a a regular RR interval. The bottom panel illustrates sinus bradycardia with a heart rate of less than 60 beats per minute, character section of the image displays a detailed ECG reading with multiple leads labeled aVR, aVL, aVF, V1, V2, V3, V4, V5, and V6, indica





(a) a description of a SB abnormality



(b) abnormality present

(5) *Sinus tachycardia (ST)* [31], described in Fig. 8, is the converse: a sinus rhythm with an elevated rate (resting heart rate > 100 bpm features (one P wave before each QRS, normal P wave axis), but the R–R intervals are shortened [32]. At very high rates, P waves may become difficult to distinguish, but each complex still has a discrete P wave onset. Typically, lead II shows the regular narrow QRS com tachycardia is most often a normal physiological response (exercise, fever, anxiety) or a result of pathology (hypovolemia, infection, hyp pathologic tachyarrhythmias by the presence of consistent P waves. In ML-based analysis, identifying sinus tachycardia involves detec morphology. If P waves become obscured, alternate leads or rhythm context are used to confirm that the rhythm remains sinus. A com

Fig. 8

Sinus tachycardia in which P waves are hidden with each preceding T-wave

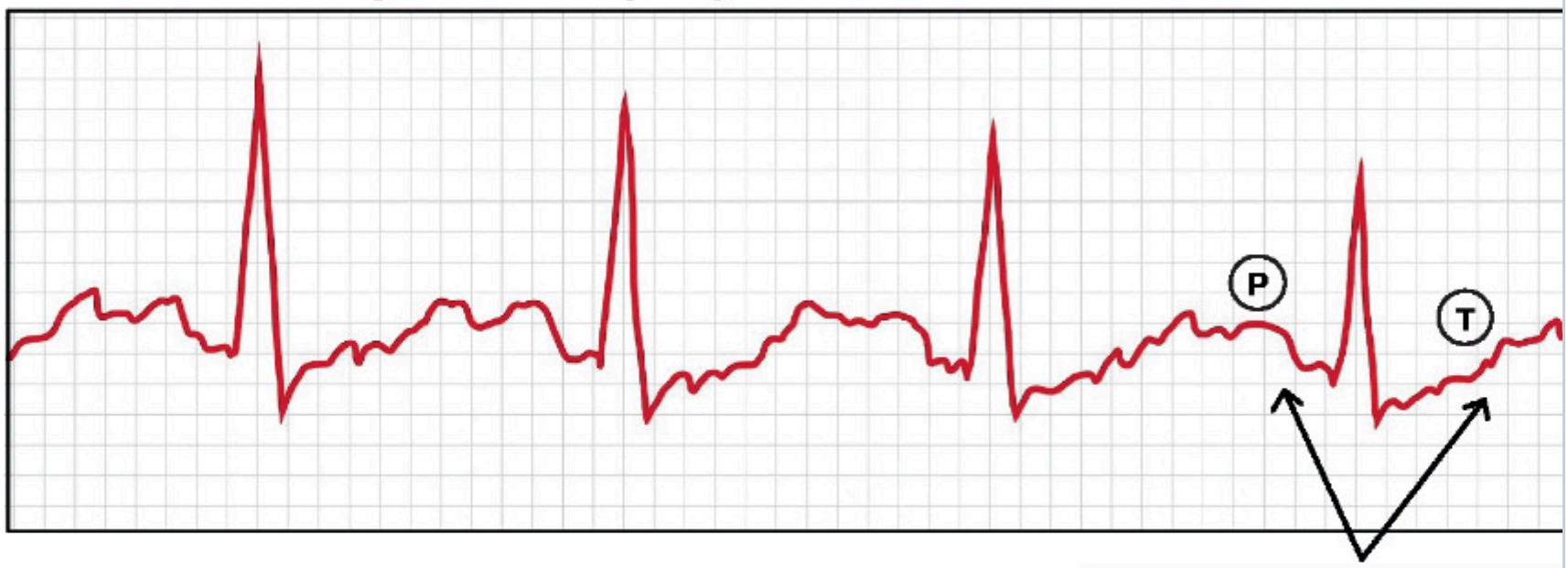
Alt Text

The image consists of two panels comparing electrocardiogram (ECG) signals. The top panel shows a "Normal Sinus Rhythm" wi
P, QRS, and T waves. The bottom panel illustrates "Sinus Tachycardia (ST)" with a heart rate over 100 bpm, highlighting a shorte
include annotations pointing to specific waveforms. Below, there is a detailed ECG strip labeled with descriptions of ST abnorma

Normal Sinus Rhythm

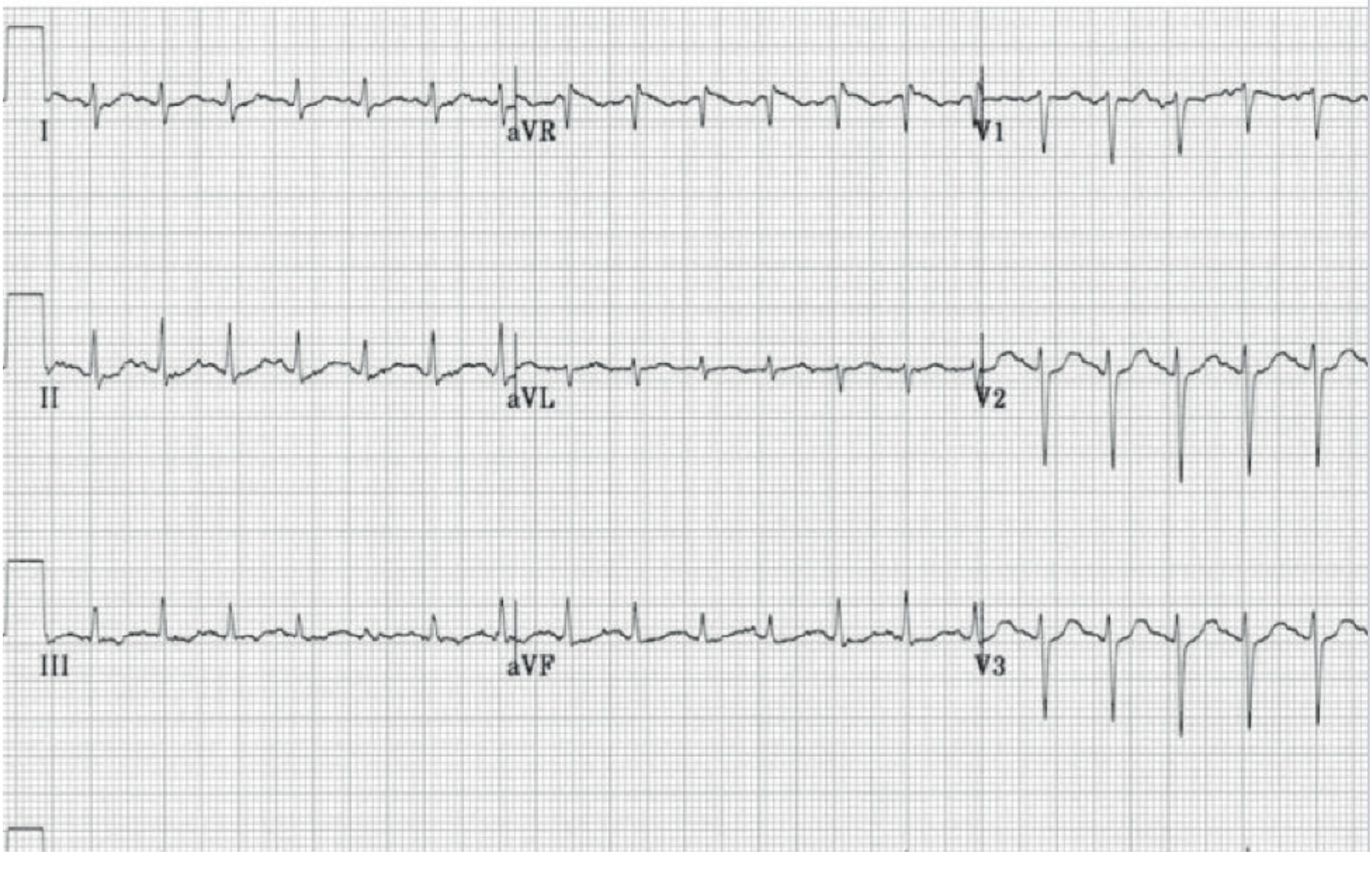
Distinguishable P and T wav

Sinus Tachycardia (ST)



Undistinguishable, merg
P and T waves

(a) description of a ST abnormality



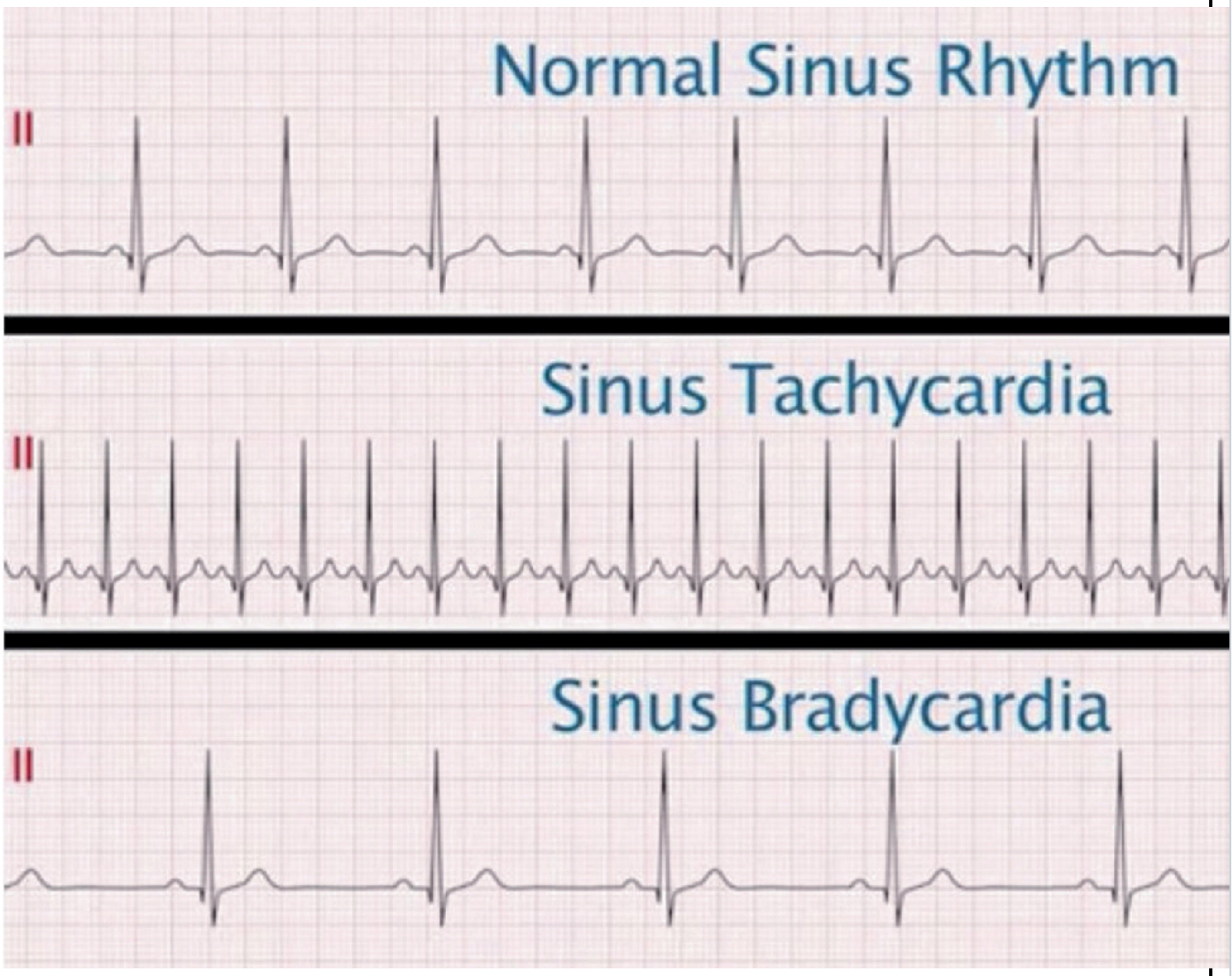
(b) abnormality present

Fig. 9

A comparison of sinus bradycardia and sinus tachycardia

Alt Text

Three electrocardiogram (ECG) tracings illustrating different heart rhythms. The top tracing shows a normal sinus rhythm with ev
sinus tachycardia with closely spaced beats, indicating a faster heart rate. The bottom tracing depicts sinus bradycardia with wid
Each tracing is labeled accordingly.



(6) *Atrial fibrillation (AF)* [33], described in Fig. 10, is the most common sustained cardiac arrhythmia. It is characterized by disorganized response. In AF, chaotic atrial depolarizations (often originating from pulmonary vein triggers or multiple micro-reentrant circuits) bomb the ventricles [34, 35]. On the ECG, this manifests as no distinct P waves and a completely irregular R-R interval. The baseline may sh isoelectric line. The QRS complexes themselves are usually narrow ($< 120ms$) unless there is a concurrent bundle branch block. Clini irregular ventricular rhythm with no visible P waves before the QRS complexes. Because AF predisposes to thromboembolic stroke and ML diagnosis, AF stands out by the absence of organized P waves and the irregular timing of QRS events. For example, deep networks pattern and lack of P waves that define AF.

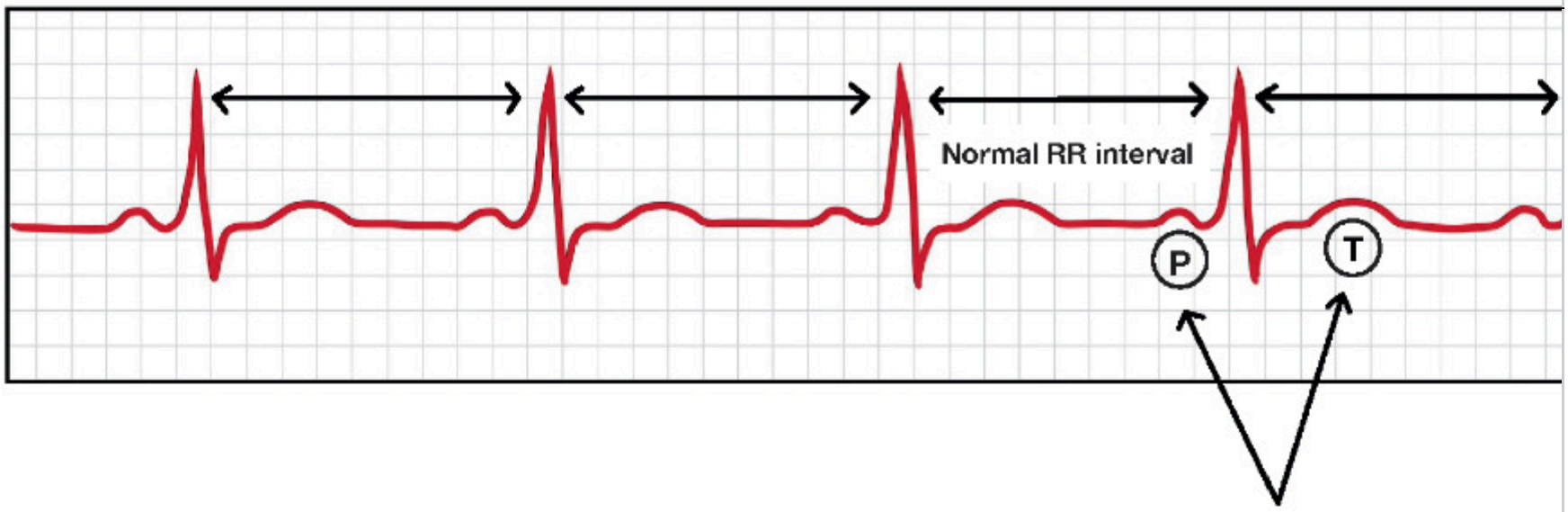
Fig. 10

Atrial fibrillation with irregular R–R intervals and absent P waves in lead II

Alt Text

The image consists of two panels illustrating electrocardiogram (ECG) signals. The top panel compares a normal sinus rhythm wi
consistent RR interval, to atrial fibrillation (AF) with indistinguishable, merging P and T waves and an inconsistent, shortened RR i
leads displaying an AF abnormality, with irregular wave patterns across different leads labeled I, II, III, aVR, aVL, aVF, V1, V2, V3, V

Normal Sinus Rhythm

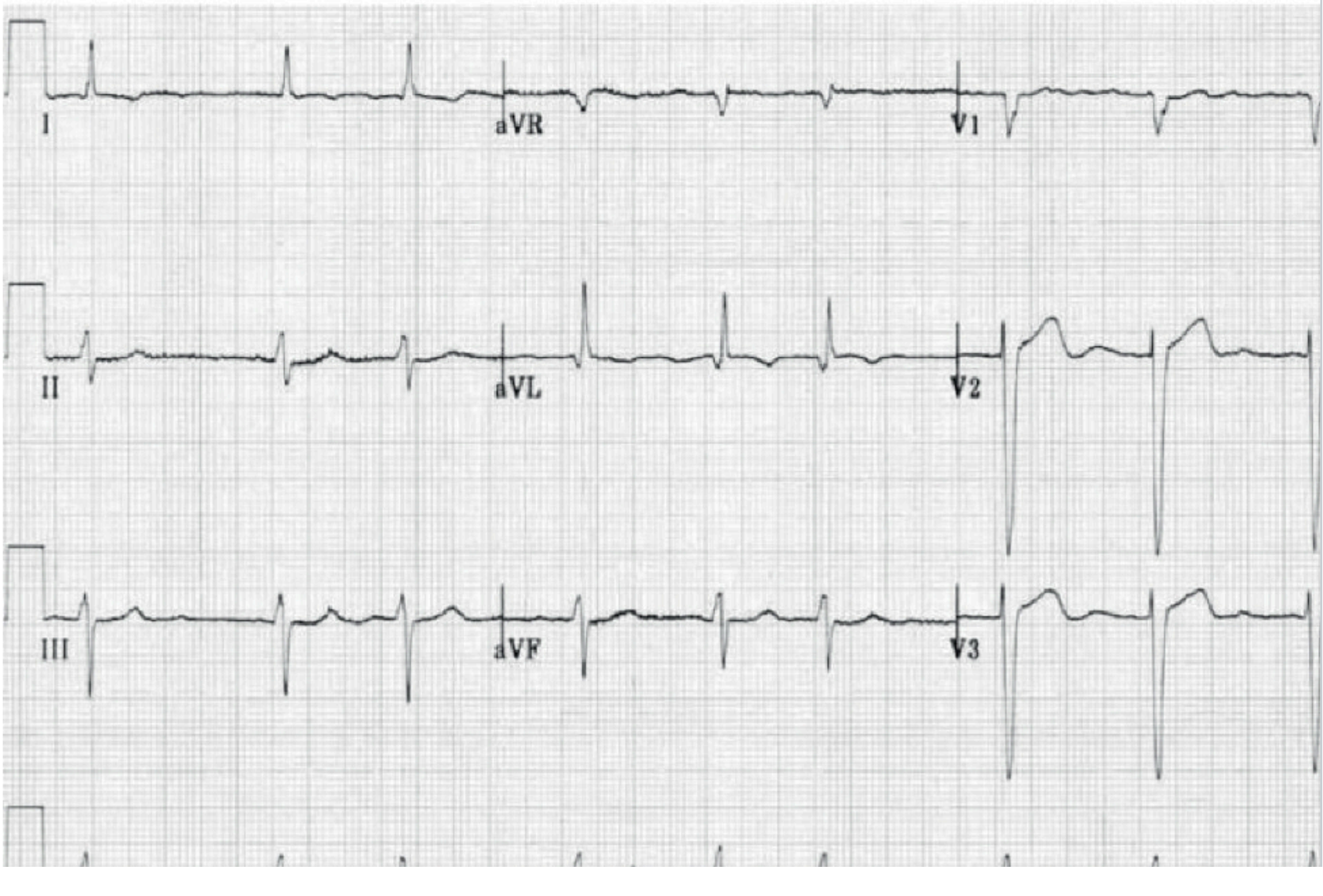


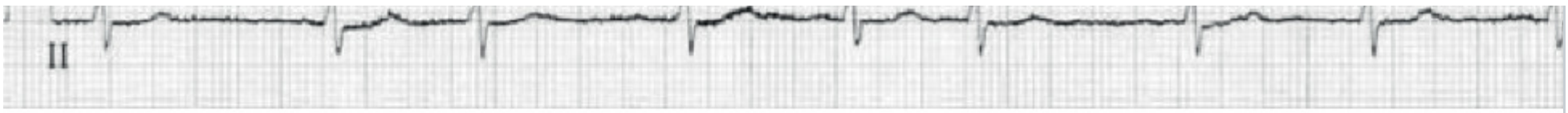
Distinguishable P and T wave

Atrial Fibrilltion (AF)

Undistinguishable, mergin
P and T waves

(a) description of an AF abnormality





(b) abnormality present

An often overlooked complication to this task is that one or more of these conditions can occur in an ECG. This kind of multi-label classification is more difficult because the training data usually does not support all combinations and permutations of the labels, making the data inherently sparse. We will discuss this in greater detail in Sect. 5.

The recent surge in artificial intelligence (AI) popularity, which stems from its success in applications such as multichannel signal and image processing, is due in part to its ability to circumvent the feature engineering problem and automatically learn complex features. Having overcome challenges like vanishing gradients and overfitting, DL in signal processing, once limited by high computational demands, delivered breakthroughs in performance using well-known networks such as Residual Network (ResNet) [36] and self-attention-based transformers [37]. Introduced as a solution to core limitations such as vanishing gradients, ResNet pioneered the use of residual connections, which allowed very deep networks to be trained efficiently, and resulted in significant gains in performance.

4. An Overview of State of the Art

The ResNet18 architecture, summarized in Figs. 11 and 12, and described in detail in [36], offers significant advantages for ECG signal processing by leveraging identity-based skip connections to learn residual functions. These connections preserve earlier feature representations and ensure stable gradient flow, enabling deeper networks to be trained without suffering from vanishing gradients. This is critical in ECG analysis, where both fine-grained waveform details (e.g., P waves and QRS complexes) and long-range temporal dependencies, such as irregular rhythms or conduction delays, must be accurately captured. The hierarchical structure of the 18-layer ResNet network allows early layers to focus on local features, while deeper layers integrate these into broader temporal patterns across cardiac cycles. Batch normalization throughout the network improves robustness against typical ECG noise, such as baseline drift and motion artifacts.

Fig. 11

A typical block in the ResNet18 architecture

Alt Text

Flow chart illustrating a multi-layered process. It begins with an input labeled "Input [256,256]" in the Input Layer. Each subsequent layer, labeled Layer 1 to Layer 3, includes operations such as "Downsample" and "DenseBlock" with varying dimensions. Arrows indicate the flow of data between these operations. The process concludes with an "Output [8,8]" in the Output Layer. Each layer is color-coded and labeled on the right.

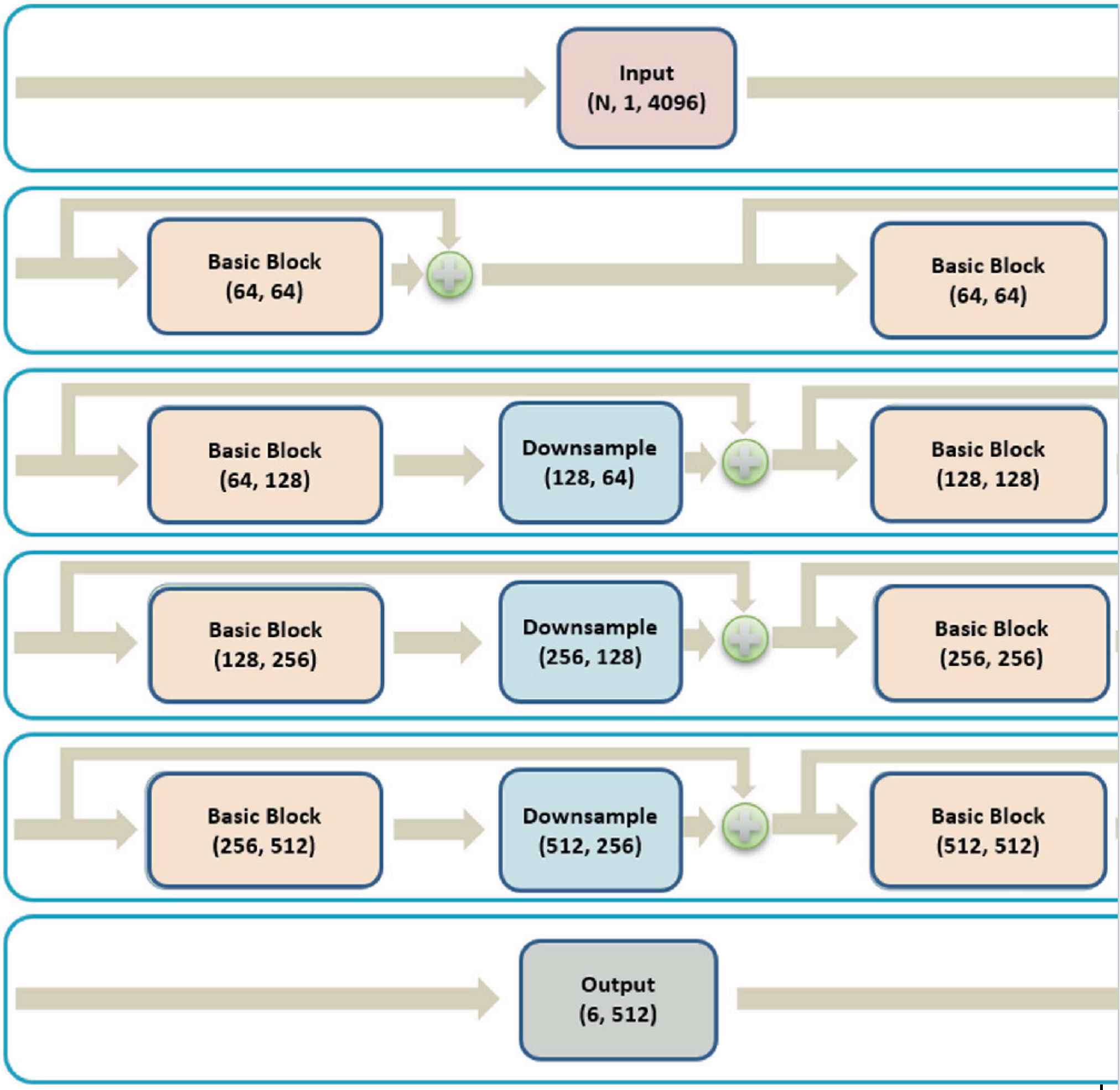
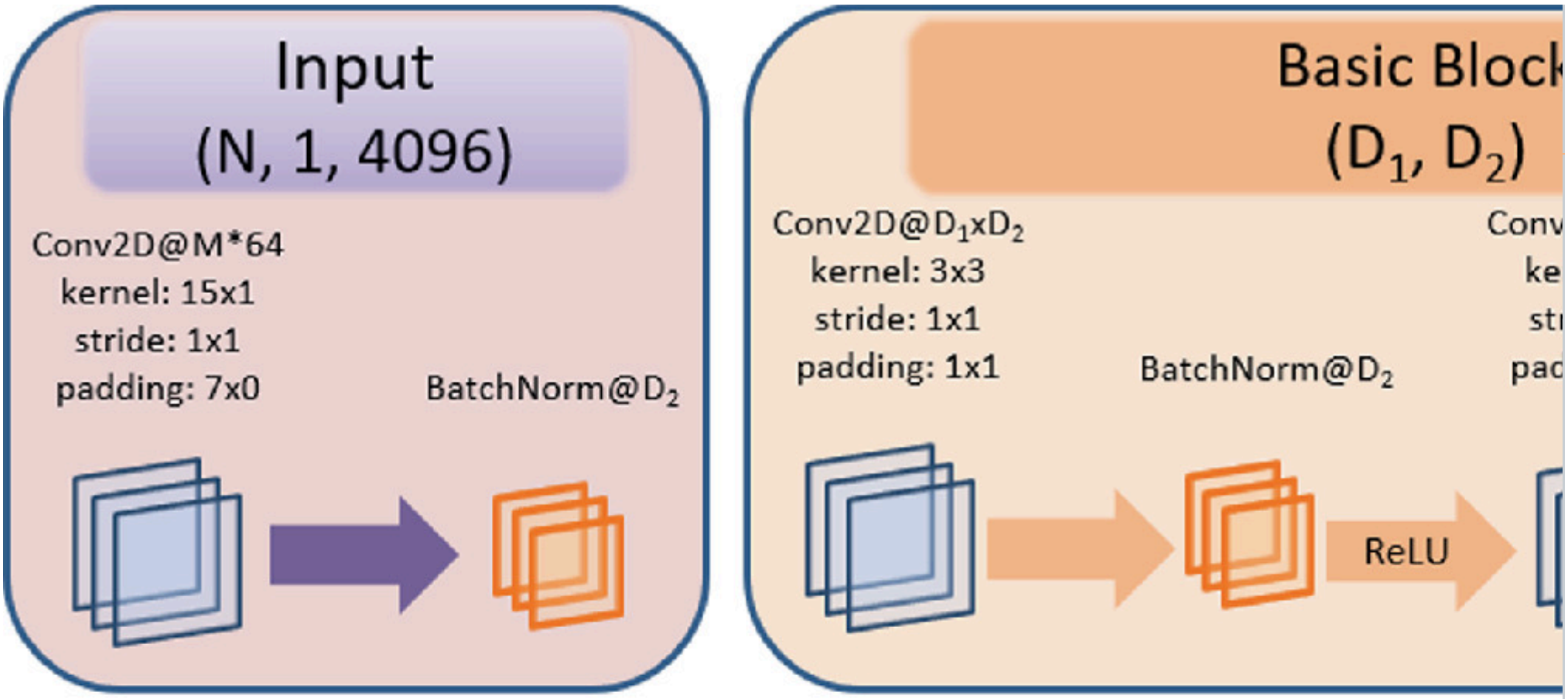


Fig. 12

The composite ResNet18 architecture uses four internal layers

Alt Text

Flow chart illustrating a neural network architecture. The process begins with an "Input" section labeled (N, 1, 4096), detailing a convolutional layer with kernel size 3x3, stride 1, and padding 1, followed by BatchNorm and ReLU activation. The "Basic Block" section, labeled (D1, D2), includes two convolutional layers with similar parameters, each followed by BatchNorm and ReLU. The "Downsample" section, labeled (D1, D2), shows a convolutional layer with kernel size 1x1, stride 1, and padding 0, followed by BatchNorm. The "Output" section, labeled (512, 6), includes a global average pooling layer and a fully connected layer. Arrows indicate the flow of data through the network.



Despite its depth, the architecture remains computationally efficient with approximately 11.7 M parameters, though this imposes constraints on the dataset size. This is a relatively modest-sized network by today's standards. State-of-the-art systems today often utilize hundreds of billions to trillions of parameters. Overall, ResNet18's use of residual blocks, convolutional layers, batch normalization, ReLU activations, and skip connections makes it well suited for ECG tasks requiring sensitivity to both local morphology and global rhythm patterns. A major goal of this chapter is to demonstrate that modern deep learning systems are capable of automatically learning the optimal relationships between ECG channels, and hence, there is no longer a need to do model-based feature generation.

Recent studies by Ribeiro et al. [17], Pastika et al. [38], Sau et al. [39], and von Bachmann et al. [40] demonstrate that deep learning models can process raw ECG signals effectively without relying on derived channels or extensive preprocessing. These results raise the question of whether all recorded channels are necessary for accurate automated diagnosis. Rather than focusing solely on reducing the number of input channels, this chapter investigates whether models trained on reduced channel configurations can outperform or match those using the full 12-channel set. Experimental results are presented to evaluate performance across varying dataset sizes and lead configurations, isolating the diagnostic value of each channel group. The analysis emphasizes model accuracy, signal redundancy, and computational efficiency, establishing whether reduced configurations provide practical and statistical advantages in multi-label cardiac classification tasks.

The k-Nearest Neighbors (KNN) algorithm proved effective for initial attempts at automated diagnosis, achieving accuracy rates between 95 and 97% [41] in identifying basic arrhythmias. Support Vector Machines (SVM) demonstrated particular utility in handling the complex, non-linear nature of ECG signals, while Random Forest approaches offered robust performance in dealing with noise and signal variations [42, 43]. Early applications focused primarily on single-lead ECG interpretation, with researchers using feature extraction techniques to identify key signal characteristics. These features included QRS complex morphology, RR intervals (the length of a ventricular cardiac cycle), and various time-domain and frequency-domain measurements. The process of feature

selection proved crucial, as the high-dimensional nature of ECG signals required careful consideration of which characteristics would provide the most diagnostic value.

A significant advancement came with the development of ensemble methods. By combining multiple algorithms, researchers achieved improved robustness and accuracy. Gradient boosting and AdaBoost techniques demonstrated success in handling the inherent variability of ECG signals, with accuracy rates between 85 and 95% for common cardiac abnormalities [44]. The challenge of preprocessing remained significant throughout this period. Traditional machine learning approaches required careful signal conditioning, including noise removal, baseline wander correction, and signal normalization. The preprocessing step proved both critical and limiting, as the quality of the analysis depended heavily on the effectiveness of these preliminary transformations.

Traditional models relied heavily on hand-crafted features, requiring significant domain expertise. Many early approaches struggled with large datasets and multiple lead configurations. The need for extensive signal preprocessing introduced potential points of failure and reduced generalizability. By the mid-2010s, researchers began exploring more sophisticated approaches. The introduction of neural networks marked a transition point, though early implementations still relied heavily on feature engineering. These initial neural network applications demonstrated promise in handling the temporal nature of ECG signals but were limited by available computational resources and dataset sizes [45]. The evolution toward deep learning approaches began to address many of these limitations. Convolutional Neural Networks (CNNs) showed promising results in automatically learning relevant features from raw ECG signals [46]. This capability reduced the dependence on manual feature engineering and improved the scalability of automated analysis systems.

Several key studies demonstrated the potential of neural network approaches during this transition period. Kiranyaz et al. [47] demonstrated the effectiveness of 1D CNNs for patient-specific ECG classification. Acharya et al. [46] showed that deep learning could effectively identify different types of arrhythmias. Hannun et al. [48] established that deep learning could match or exceed human performance in single-lead ECG interpretation. Long Short-Term Memory (LSTM) [49] networks became a powerful tool for classifying electrocardiogram (ECG) and cardiovascular data. Several studies [50, 51, 52] have explored their application in ECG classification, as these recurrent neural networks have demonstrated good performance in capturing temporal dependencies in sequential data. Despite their success, LSTM networks face certain challenges in ECG classification. One challenge is the need for large amounts of labeled data for training, which can be difficult to get in medical applications. Also, the computational requirements of LSTM networks may limit their use in resource-limited environments, although efforts have been made to develop lightweight versions for wearable devices [50, 51, 52, 53].

Transformer-based models [54] have shown remarkable success in various domains, including natural language processing, computer vision, and time series forecasting. However, their application to ECG classification presents unique challenges, particularly when dealing with long-term sequential data [55]. Unlike discrete and highly semantic language tokens, ECG data consists of continuous numeric points with temporal redundancy and weak semantics. This characteristic makes it challenging for transformers to effectively depict the overall properties of time series, such as trends and periodic variations [56]. Such systems often require large amounts of training data and are extremely sensitive to imbalanced data. Despite the enormous size of the TNMG Corpus, the underrepresentation of certain combinations of labels poses significant challenges for transformer-based models. Hence, we selected a ResNet architecture that has performed fairly reliably across a wide range of applications.

The groundwork laid by these early machine learning applications proved essential for later developments in deep learning approaches. The challenges encountered and solutions developed during this period informed the design of more sophisticated systems, leading ultimately to comprehensive approaches such as those developed by Ribeiro et al. [17]. One significant development was the recognition that different types of cardiac abnormalities might require different analytical approaches. While some conditions could be identified through relatively simple feature analysis, others required more sophisticated pattern recognition capabilities. This understanding influenced the development of hybrid approaches that combined multiple analytical techniques.

The field continued to evolve with the introduction of transfer learning techniques, which allowed models trained on large datasets to be effectively adapted for more specific applications. This approach proved particularly valuable in addressing the challenge of limited data availability for rare cardiac conditions, such as Brugada syndrome, Wolff-Parkinson-White syndrome, and certain types of congenital long QT syndromes. These conditions are underrepresented in most clinical datasets, making it difficult to train accurate models directly. Transfer learning enables the reuse of generalized low-level ECG features learned from common conditions and adapts them to identify rare patterns through fine-tuning on small, condition-specific datasets [57, 58]. This strategy increases model robustness and diagnostic reach without requiring extensive new data collection.

The field has subsequently established several foundational principles for effective ECG analysis. The importance of end-to-end learning approaches that could process raw ECG signals directly became evident [59, 60]. An architecture capable of handling multiple lead configurations effectively proved crucial [17]. The need for robust validation approaches using clinically relevant metrics emerged as a key consideration [61]. Additionally, the importance of interpretability in automated analysis systems became increasingly recognized [62].

These developments set the stage for more comprehensive approaches to automated ECG analysis, leading to systems capable of handling multiple lead configurations and identifying a broader range of cardiac abnormalities. The progression from traditional machine learning to deep learning approaches marked a significant advancement in the field's capability to provide reliable, automated cardiac diagnosis support. Most systems today use transfer learning, fine-tuning, and reinforcement learning to achieve high performance. We are in the early stages of such work [63].

5. Baseline Performance on TNMG Code

The TNMG CODE Corpus (TNMG) [16] represents a major advancement in the field of cardiology and is the corpus we focus on in this study. TNMG is a dataset of ECG records collected by the Telehealth Network of Minas Gerais (TNMG) between **2010** and **2016** in **811** counties in the Brazilian state of Minas Gerais, organized by the Clinical Outcomes in Digital Electrocardiography (CODE) group. The dataset contains a total of **6,716,317** annotated records from **1,558,749** patients.

The dataset includes a curated "golden dataset" of **827** ECG recordings, which serves as a high-quality evaluation set. These recordings were independently annotated by two cardiologists. In cases of disagreement, a third specialist reviewed the annotations to establish a consensus. The dataset was labeled for six abnormalities discussed in Sect. 2 and shown in Table 1.

Table 1

Annotations present in TNMG

Label	Description
1dAVb	First-degree atrioventricular block: A delay in the conduction of electrical impulses from the atria to the ventricles
RBBB	Right bundle branch block: A condition where the right side of the heart's electrical conduction system is impaired
LBBB	Left bundle branch block: A condition where the left side of the heart's electrical conduction system is impaired
SB	Sinus bradycardia: A slower-than-normal heart rhythm, defined as a heart rate below 60 beats per minute in adults
AF	Atrial fibrillation: An irregular heart rhythm, characterized by chaotic electrical activity
ST	Sinus tachycardia: A higher-than-normal heart rhythm, defined as a heart rate above 100 beats per minute in adults

In Table 2, we show a distribution in the number of records and percentage of feature vectors in both datasets, where the presence or absence of each abnormality is marked as a binary vector in the same order as Table 1. It is clear that the majority of both datasets consists of healthy records. The imbalance in this data has a profound impact on our ability to train high-performance models. However, it is not our intention here to focus on techniques to deal with imbalance. Tokens with a single disease occur in single-digit percentages. An even smaller fraction of tokens with multiple diseases appears in the corpus. Of equal concern is that tokens with multiple diseases are not well represented in the evaluation dataset, known as the gold standard dataset. Hence, despite the enormous amount of data present in TNMG, the sparsity of the data does pose challenges, especially for multi-label classification problems. These can be addressed using any of the popular techniques for dealing with imbalance [64, 65, 66].

Table 2

Distribution of classes in TNMG CODE

Feature vector	Train		Gold (eval)	
	#	%	#	%
000,000	6,014,462	89.55000	681	82.34583
010,000	145,208	2.16202	28	3.38573
000,001	131,820	1.96268	35	4.23216
000,010	100,865	1.50179	11	1.33011
000,100	94,500	1.40702	15	1.81378
001,000	86,487	1.28771	20	2.41838
100,000	75,924	1.13044	25	3.02297
010,010	11,910	0.17733	1	0.12092
110,000	11,168	0.16628	0	0.00000
101,000	7,580	0.11286	3	0.36276
001,010	7,019	0.10451	0	0.00000
010,100	5,713	0.08506	0	0.00000
100,100	4,215	0.06276	0	0.00000
010,001	3,408	0.05074	1	0.12092
001,001	3,066	0.04565	0	0.00000
000,011	2,860	0.04258	0	0.00000
100,010	1,871	0.02786	1	0.12092
001,100	1,625	0.02419	1	0.12092
011,000	1,621	0.02414	4	0.48368
110,100	1,165	0.01735	0	0.00000
100,001	560	0.00834	1	0.12092
110,010	515	0.00767	0	0.00000

Feature vector	Train		Gold (eval)	
	#	%	#	%
000,110	506	0.00753	0	0.00000
101,100	331	0.00493	0	0.00000
011,100	329	0.00490	0	0.00000
010,011	292	0.00435	0	0.00000
101,010	248	0.00369	0	0.00000
111,000	220	0.00328	0	0.00000
011,010	189	0.00281	0	0.00000
001,011	143	0.00213	0	0.00000
100,110	88	0.00131	0	0.00000
010,110	82	0.00122	0	0.00000
110,110	64	0.00095	0	0.00000
111,100	56	0.00083	0	0.00000
011,001	45	0.00067	0	0.00000
110,001	43	0.00064	0	0.00000
100,011	36	0.00054	0	0.00000
111,010	19	0.00028	0	0.00000
101,001	16	0.00024	0	0.00000
001,110	14	0.00021	0	0.00000
000,101	10	0.00015	0	0.00000
011,011	5	0.00007	0	0.00000
010,101	5	0.00007	0	0.00000
011,110	5	0.00007	0	0.00000
111,011	4	0.00006	0	0.00000
111,110	3	0.00004	0	0.00000
110,011	1	0.00001	0	0.00000
101,110	1	0.00001	0	0.00000

The application of deep neural networks (DNNs) to ECG analysis represents a significant advancement in automated cardiac diagnosis. The study conducted by Ribeiro et al. [17] demonstrates both the potential and limitations of this approach on the TNMG CODE corpus. The researchers employed a residual network architecture (ResNet18) adapted for unidimensional signals. Their unique and careful approach to data labeling combined automatic analysis using the University of Glasgow ECG analysis program with expert cardiologist annotations, implementing a systematic procedure to resolve discrepancies. This dual-verification approach strengthens the reliability of their training data. We chose this study because it is one of the few large-scale, real-world uses of deep learning for ECG classification with open code and full dataset access. This made it a strong base for our work, with both solid methods and practical use.

For preprocessing, they resampled all ECGs to $400\mathrm{Hz}$, zero-padded signals to 4096 samples per lead, used the derived channels, and applied z-score normalization. While their approach showed promising results, the impact of this preprocessing on model performance was not thoroughly investigated. Their decision to make all data the same length somewhat simplifies the sequential decoding nature of the problem but also makes issues such as normalization easier to resolve. Results are not significantly sensitive to issues such as the normalized sample frequency and methods of interpolation.

Newer studies by Pastika et al. [38] and von Bachmann et al. [40] have also adopted reduced lead configurations, utilizing 8-lead ECGs in their deep learning models for body mass index and electrolyte prediction, respectively. However, these studies did not extensively discuss the rationale for lead reduction. The choice of using raw leads stems from the fact that these additional leads are linear combinations of the raw leads, making them redundant in a deep learning context.

Following Ribeiro et al. [17], we utilize a ResNet18 architecture adapted for multi-label classification. To create input tensors for our model, we transform the time series ECG data into multichannel 3D tensors. Following [17], all signals are zero-padded to 4096 samples and undergo z-score normalization to standardize the data. The normalized signals are then reshaped into tensors of shape $(N, 1, 4096)$, where N represents the number of ECG channels. In this representation, each channel corresponds to a separate ECG lead, and the temporal samples span the width of the tensor. Although the tensor has three dimensions, the singleton height dimension (1) facilitates the processing of temporal data similarly to how CNNs handle spatial information in images.

A typical block in our architecture is shown in Fig. 11. The overall architecture is illustrated in Fig. 12. The “basic block” shown in Fig. 11 is the fundamental building unit of the system [36]. It has two convolutional layers, each followed by batch normalization and a ReLU nonlinearity, plus a shortcut (skip connection). We made minor modifications to the original architecture. The first convolutional layer of our model was modified to take either eight or twelve channels as input. The final layer was adapted to output

probabilities for each of the six cardiac conditions by utilizing a sigmoid activation function. We employ the Adam optimization algorithm [67] with a learning rate of 0.001. Due to the multi-label nature of our task, we use Binary Cross-Entropy loss [68] as the objective function:

$$BCE\,(y, \hat{y}) \;=\; -\frac{1}{n}\sum\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right] \tag{5}$$

where *n* is the number of classes, y_i is the true label, and $\hat{y}_i$ is the predicted probability for class *i*.

In all experiments, the training process iterated over 10 epochs with a batch size of 32. We monitored training and validation losses, along with accuracy, micro-averaged F1 score, and macro-averaged F1 score to assess the model performance. This approach allows us to evaluate the effectiveness of our model in processing both raw and minimally preprocessed ECG signals.

Given that our problem is a multi-label classification task, we employ micro and macro F1 scores [69] as a key metric for evaluating model performance. These metrics provide a comprehensive assessment of the overall model's ability to identify several cardiac conditions at once. The micro F1 score calculates metrics by counting the true positives, false negatives, and false positives across all classes. It is computed as the harmonic mean of precision and recall:

$$\mathbf{Micro\ F1} \;=\; 2 * \frac{(\mathbf{Precision} * \mathbf{Recall})}{(\mathbf{Precision} + \mathbf{Recall})}. \tag{6}$$

$$\mathbf{Precision} \;=\; \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FP}} \tag{7}$$

$$\mathbf{Recall} \;=\; \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FN}} \tag{8}$$

where TP, FP, and FN represent true positives, false positives, and false negatives, respectively.

In contrast, the macro F1 score calculates the F1 score for each class independently and then averages these scores. Micro F1 tends to give more weight to frequent classes, while macro F1 gives equal weight to all classes, regardless of their frequency in the dataset.

6. Channel Reduction

We conducted eight experiments to systematically evaluate the impact of ECG channel reduction and dataset size on deep learning performance. We used four different training dataset sizes: $\mathbf{2K\,(2,000)}$, $\mathbf{20\,K\ \ (20,000)}$, $\mathbf{200K\,(200,000)}$, and $\mathbf{2,000K\,(2,000,000)}$ records, each tested with the 8-channel and 12-channel ECG configuration. For the 12-channel sets, we applied minimal preprocessing consisting of resampling to $\mathbf{400Hz}$ and $2x$ scaling of the signal amplitude. To address class imbalance, we chose the distribution for each training set that balances the frequency of occurrence of the class labels to the extent possible. However, as the dataset size increased, we had to include a higher proportion of healthy records due to their prevalence in the corpus.

In each experiment, we trained a separate ResNet18 model and evaluated its performance on the golden test set. We also utilized a fixed development set of 5,000 records in all experiments to monitor the training process and prevent overfitting. The development set was balanced to represent a variety of combinations of cardiac conditions: approximately 4,000, examples were evenly split between single-condition cases and healthy records, 750 evenly split examples with two conditions, 249 examples containing three conditions, and one rare example with four concurrent conditions. The results for each experiment are shown in Table 3.

Table 3

Micro F1 scores as a function of the training set size

Train size (K)	No. chans	Train	Dev	Eval
2	8	0.8810	0.7024	0.5029
2	12	0.8690	0.7050	0.2127
20	8	0.8870	0.8288	0.7022
20	12	0.8812	0.8366	0.5509
200	8	0.9310	0.8461	0.8421
200	12	0.9286	0.8545	0.7956
2,000	8	0.8809	0.7787	0.8649
2,000	12	0.8787	0.7708	0.8522

A breakdown of the errors by label is given in Table 4. Our results reveal a consistent pattern across all dataset sizes: the models trained on 8-channel ECG data outperformed those trained on 12-channel data. The performance difference was more pronounced in the smaller datasets and gradually diminished as the training dataset size increased.

Table 4

A breakdown of errors on the eval set by label as a function of the amount of training data

Label	Freq	12 Channels (derived)			8 Channels (raw)		
		2K	200K	2M	2K	200K	2M
IdAVb	28	24	17	18	22	11	14
RBBB	34	27	6	5	64	5	6
LBBB	30	22	2	2	13	1	2
SB	16	51	26	8	49	22	8
AF	13	365	11	1	41	5	2
ST	37	44	13	9	70	13	8

We observed a significant decline in performance in both 8- and 12-channel models in experiments with $2,000\mathrm{K}$ records. Performance on the training set, which constitutes closed-set testing, declined in going from $200\mathrm{K}$ records to $2,000\mathrm{K}$ records. Performance on the development set similarly declined. However, performance on the blind evaluation set improved slightly. We attribute this decrease to the inherent class imbalance in the larger dataset. As we expanded to a much higher number of records, the proportion of healthy ECG examples increased significantly. Although it reflects the prevalence of these records in the general population, this imbalance led to a bias in the model's predictions, favoring the majority class at the expense of less common combinations of cardiac conditions.

Another observation is that models trained on $2,000\mathrm{K}$ records performed poorly on the balanced development set but showed a noticeably higher performance on the evaluation set. This discrepancy is likely caused by the higher proportion of healthy records in the evaluation set, which more closely mirrors the distribution in the training data. These observations are an example of the importance of considering dataset composition and carefully balancing class distributions within datasets.

To assess the stability and reproducibility of our findings, we conducted several experiments to estimate the variance of the F1 scores on 8-channel data. For each of the three dataset sizes ($2\mathrm{K}$, $20\mathrm{K}$ and $200\mathrm{K}$) we performed five independent training runs. Each run utilized a different random seed for data shuffling and model initialization. The results are shown in Table 5. As expected, there is a significant reduction in the standard deviation as the training set size increases.

Table 5

Micro F1 scores as a function of the random seed

Data	Train	Dev	Eval
2K 8 Channels (1)	0.8898	0.7235	0.4596
2K 8 Channels (2)	0.8747	0.7047	0.4569
2K 8 Channels (3)	0.8766	0.7074	0.5111
2K 8 Channels (4)	0.8892	0.7163	0.4426
2K 8 Channels (5)	0.8720	0.6876	0.5251
StDev	0.0084	0.0136	0.0366
20K 8 Channels (1)	0.8887	0.8187	0.7335
20K 8 Channels (2)	0.8852	0.8132	0.7214
20K 8 Channels (3)	0.8880	0.8229	0.6897
20K 8 Channels (4)	0.8884	0.8241	0.6729
20K 8 Channels (5)	0.8882	0.8243	0.6905
StDev	0.0014	0.0047	0.0249
200K 8 Channels (1)	0.9312	0.8534	0.8251
200K 8 Channels (2)	0.9298	0.8523	0.8278
200K 8 Channels (3)	0.9307	0.8510	0.7831
200K 8 Channels (4)	0.9318	0.8451	0.8278
200K 8 Channels (5)	0.9298	0.8481	0.7752
StDev	0.0009	0.0034	0.0263

This translates to an improvement in the statistical significance of these scores. For a sample size of $2\mathrm{K}$ at 95% confidence, a difference in the F1 score of 0.0174 is statistically significant on the training data. For sample sizes of $20\mathrm{K}$ and $200\mathrm{K}$, differences greater than 0.0054 and 0.0017, respectively, are significant. Hence, we see that the differences due to randomization in Table 4 are statistically significant, underscoring how sensitive these deep learning systems are to randomization (which makes reproducibility a challenge).

Nevertheless, the key point here is that performance for the 8-channel system is not statistically different from the 12-channel system, indicating that the deep learning system is able to implement whatever signal processing is necessary to extract meaningful information.

The research achieved impressive performance metrics, with F1 scores exceeding 80% and specificity above 99% across the six targeted conditions. The deep neural network matched or surpassed the diagnostic performance of cardiology residents. These results highlight the method's effectiveness for automated detection of common rhythm and conduction abnormalities. However, the study's scope remains narrow. It focuses on a limited set of conditions excluding many clinically relevant pathologies. The use of data exclusively from Brazilian healthcare settings also raises concerns about generalizability to broader populations. Additionally, the classification framework models each abnormality in isolation, without accounting for diagnostic interactions or comorbid patterns often encountered in real-world ECG interpretation. These limitations highlight the need to integrate such methods into the broader historical progression of ECG analysis and to evaluate their role alongside evolving modeling strategies.

A notable limitation emerges in the handling of borderline cases. The error analysis reveals that most DNN mistakes occurred in cases involving ECG interval measurements near diagnostic thresholds. This suggests potential limitations in the network's ability to handle cases that fall in diagnostic boundary zones—a crucial consideration for clinical implementation. The choice to use the ResNet18 architecture with a quarter of the layers used in previous studies warrants more detailed justification.

7. Feature Importance

While previous research established the effectiveness of deep learning for ECG analysis, the necessity of derived ECG channels in these modern approaches remained unexplored. We systematically investigated this question through a series of carefully designed experiments. We conducted an ablation analysis [70] with 250 randomized iterations for each channel. This process quantified feature importance by measuring the average decrease in micro F1 score when each channel was randomized. The analysis demonstrated that precordial leads (V1-V6) consistently showed higher importance scores, while derived channels (DIII, aVR, aVL, and aVF) often showed lower or negative scores, indicating redundancy rather than contribution.

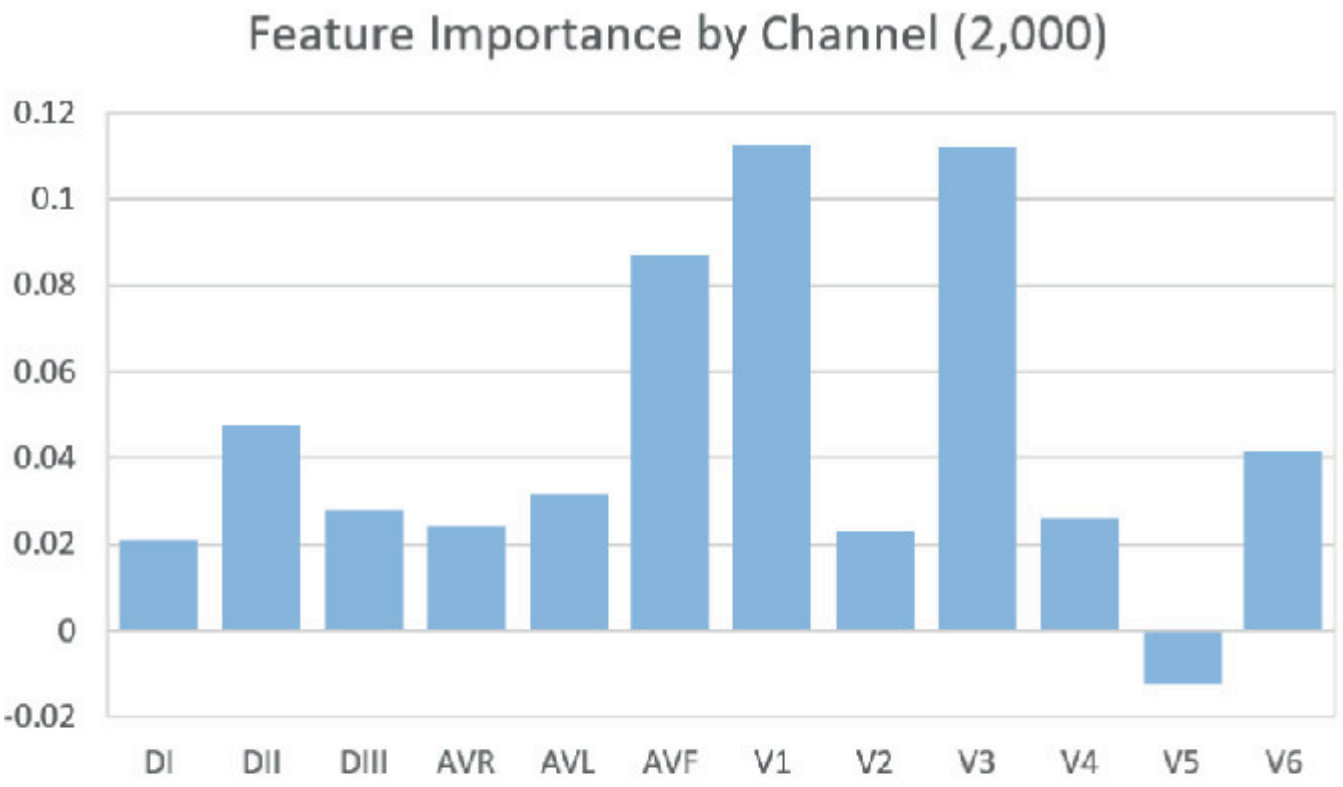
Four systematic experiments, shown in Fig. 13, were conducted to evaluate the impact of ECG channel reduction and dataset size on deep learning performance. We used four different training dataset sizes: $2K$, $20K$, $200K$, and 2,000K records, and 12-channel ECG configurations. For the 12-channel sets, minimal preprocessing was applied, consisting only of resampling to $400\ Hz$ and $2x$ scaling of signal amplitude. A critical aspect of the experimental design was the careful balancing of feature vectors in each training set to address class imbalance. However, larger datasets necessarily included more healthy records due to their prevalence in the corpus.

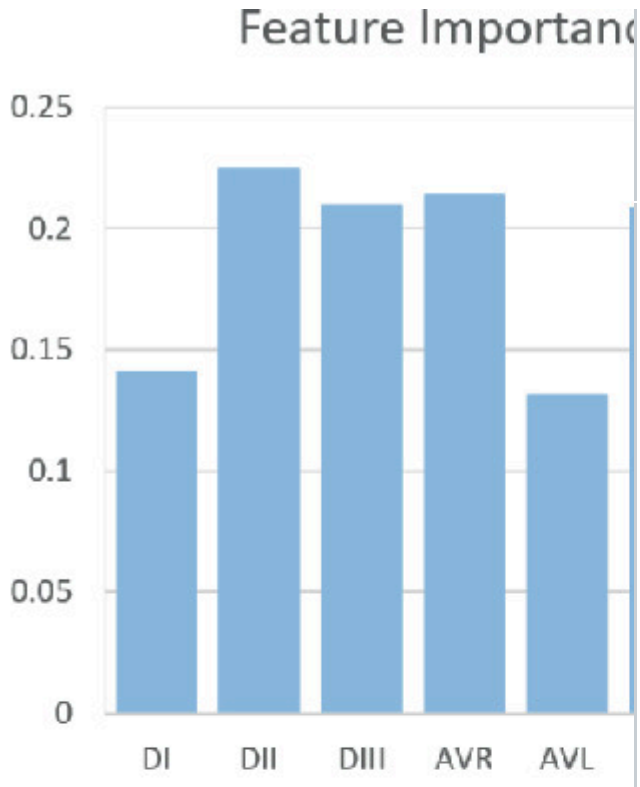
Fig. 13

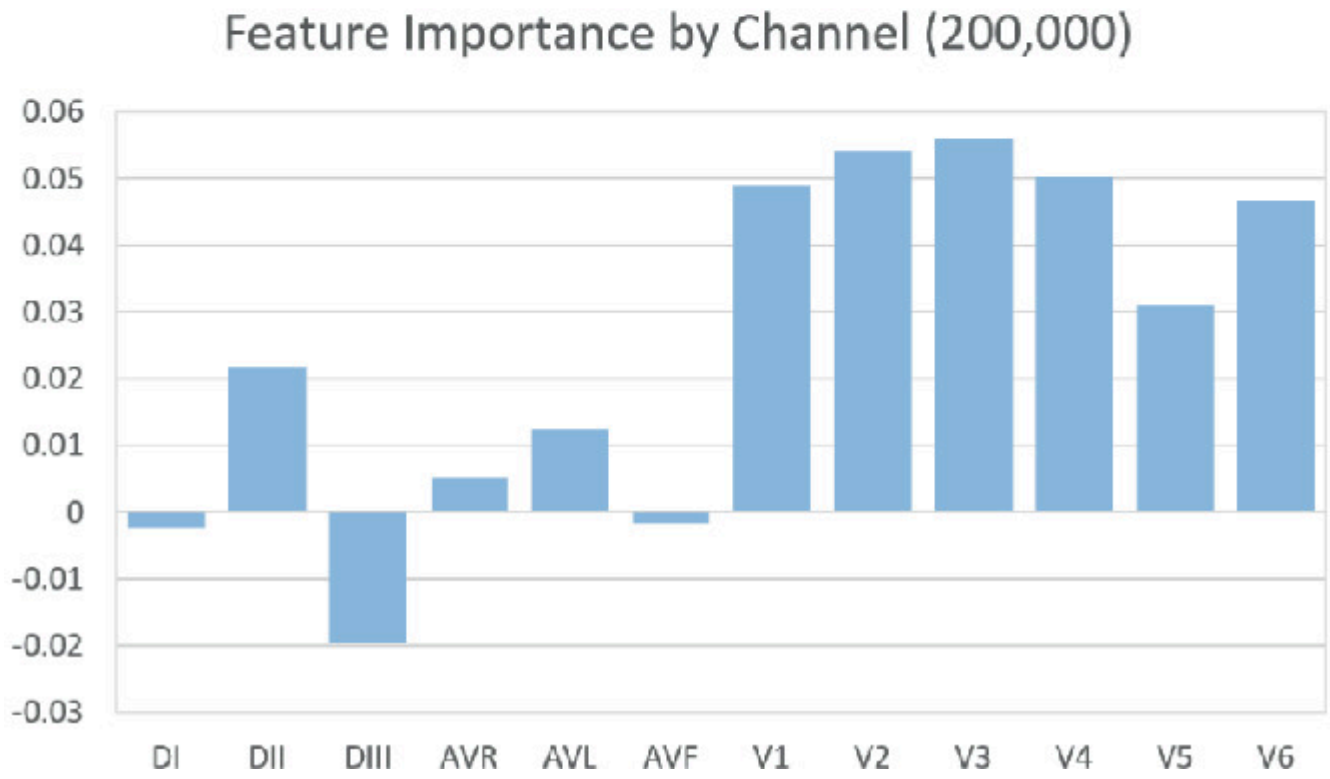
Results of the ablation analysis

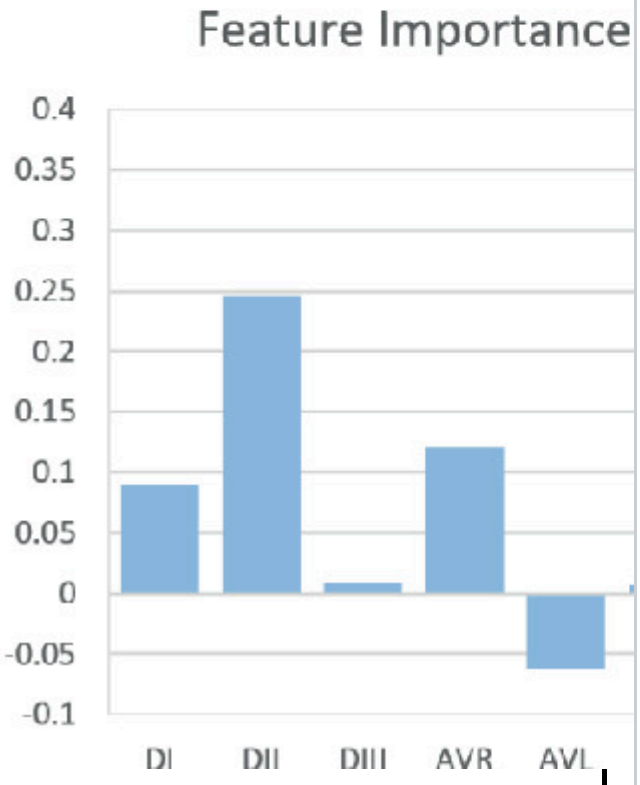
Alt Text

Bar chart figure showing "Feature Importance by Channel" across four different sample sizes: 2,000, 20,000, 200,000, and 2,000,000. Each chart displays feature importance values for channels DI, DII, DIII, AVR, AVL, AVF, V1, V2, V3, V4, V5, and V6. The heights of the bars vary, indicating the relative importance of each feature within the specified sample size.









Development and evaluation utilized three distinct datasets: the training set, a development set of 5,000 records for monitoring training progress and preventing overfitting, and a golden test set of 827 records. The development set was carefully balanced to represent various combinations of cardiac conditions: roughly 4,000 examples split evenly between single-condition cases and healthy records, 750 examples with two conditions, 249 examples containing three conditions, and one rare example with four concurrent conditions.

The experimental results revealed several key patterns. First, models trained on 8-channel ECG data consistently outperformed those using 12-channel configurations across all dataset sizes. This performance difference was most pronounced with smaller datasets and gradually diminished as training data increased. However, a significant decline in performance occurred in both 8-channel and 12-channel models when using 2,000K records. All models were trained using the Adam optimizer with a learning rate of 0.001, batch size of 32, and binary cross-entropy loss. Each experiment used five training runs with varying random seeds. Models were trained for 50 epochs with input signals resampled to a fixed length of 2,048 samples and normalized per channel. Data loaders used four workers, and training was conducted on GPU hardware.

This unexpected performance decline with the largest dataset emerged from inherent class imbalance. As the dataset expanded to include a much higher number of records, the proportion of healthy ECG examples increased significantly. Although this reflects real-world prevalence, it created a bias in the model's predictions, favoring majority class identification at the expense of accurately detecting fewer common combinations of cardiac conditions (class imbalance).

An interesting observation emerged when comparing performance on different evaluation sets. Models trained on 2,000K records showed poor performance on the balanced development set but notably higher performance on the evaluation set. This discrepancy may arise from the evaluation set's distribution more closely aligning with the training data's bias toward healthy records.

We also addressed multicollinearity in the ECG signals, shown in Fig. 14. Multicollinearity [71] occurs when there is an approximately linear relationship between two or more independent variables in a regression model. While it is typically discussed in the context of regression models, it can also affect classification models, including our deep learning model for ECG classification. We employed the Variance Inflation Factor (VIF) analysis [72] on a balanced subset of 20,000 ECG recordings. VIF quantifies the severity of multicollinearity in regression analysis by measuring how much the variance of a coefficient estimate is increased due to

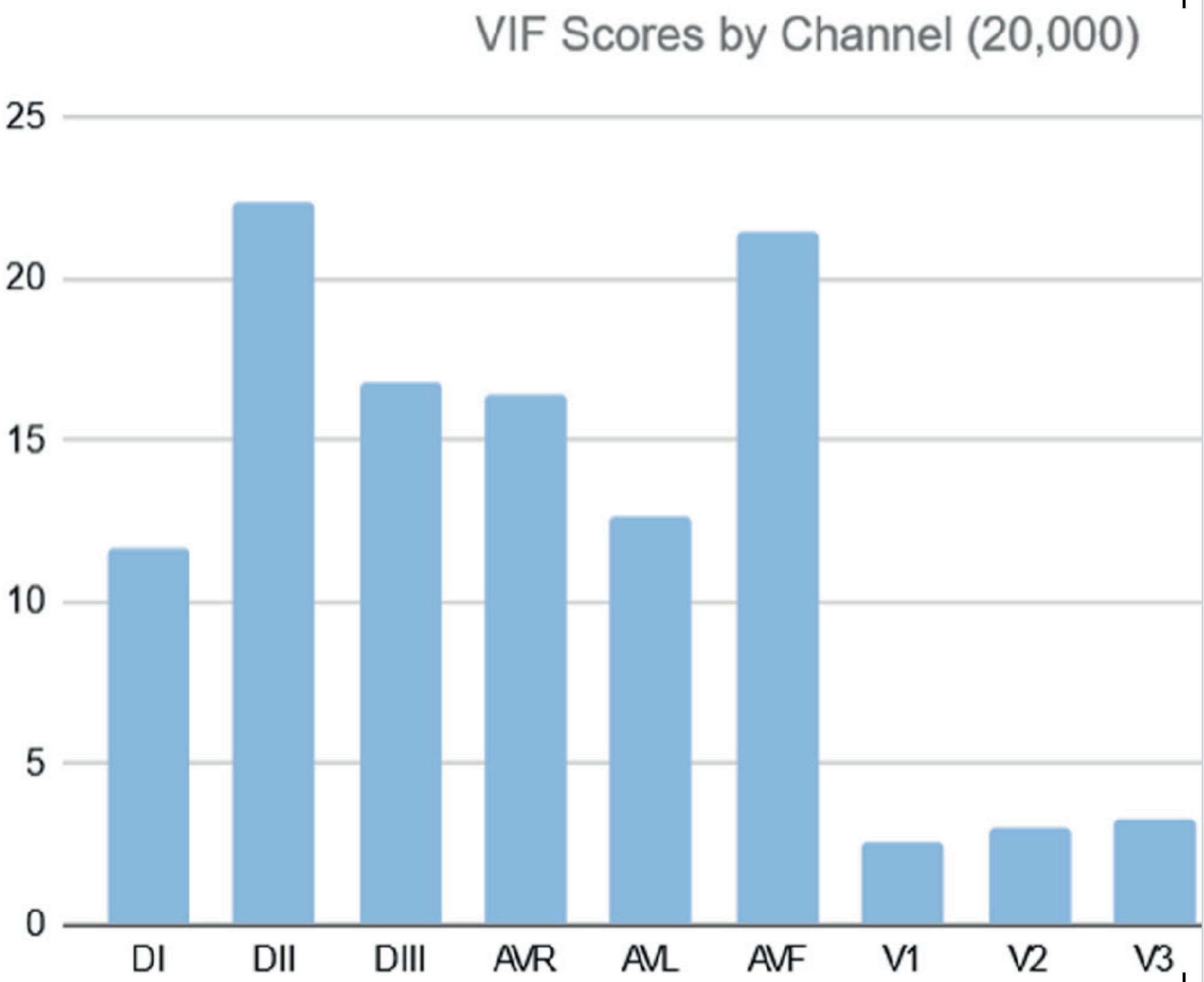
collinearity with other variables. Our analysis revealed substantial redundancy among limb leads and their derived channels, with DI and DII showing high VIF scores of 11.64 and $22.38,$ respectively. The derived channels (DIII, aVR, aVL, and aVF) similarly exhibited high VIF scores ranging from 12.69 to 21.48, indicating severe multicollinearity. In contrast, all precordial leads (V1-V6) demonstrated low VIF scores between 2.60 and 3.63, well below the standard multicollinearity threshold of 5.0. These results provide strong statistical evidence that limb leads, and their derivatives, contain redundant information, while precordial leads contribute unique and independent signal characteristics.

Fig. 14

VIF scores

Alt Text

Bar chart titled "VIF Scores by Channel (20,000)" displaying scores for different channels: DI, DII, DIII, AVR, AVL, AVF, V1, V2, V3, V4, V5, and V6. The highest scores are for channels DII and AVF, while channels V1 to V6 have lower scores. The chart uses a vertical axis for VIF scores and a horizontal axis for channel labels.



These results suggest that using only the eight independent leads effectively reduces this multicollinearity, potentially explaining the improved performance of the reduced channel configuration. This finding has important implications for signal processing in medical diagnostics more broadly. We validated the stability of these findings through additional experiments using various data splitting strategies: random 90–5–5 splits, patient-stratified splits, and chronological splits. Results showed no statistically significant difference between these approaches, except for chronological splitting, in which changes in telehealth center operations over time affected performance.

8. Conclusions

Our findings challenge traditional ECG analysis approaches and suggest that modern deep learning methods can effectively learn from raw signals without derived features. In our study, by using only the eight independent leads and omitting the derived leads, we performed a form of variable selection that addresses the issue of multicollinearity in ECG data. Our findings in model training,

ablation analysis, and VIF scores support the hypothesis, indicating that derived ECG channels introduce multicollinearity and do not provide any additional predictive power.

We point toward several future directions, including the investigation of optimal channel configurations for specific diagnostic tasks and the potential development of simplified ECG recording equipment. Our study also raises broader questions about feature engineering in medical signal processing, suggesting that traditional derived measurements may be unnecessary or detrimental when using modern machine learning approaches if there is adequate training data. Our study provides evidence supporting our hypothesis that extensive preprocessing of ECG data may not be beneficial for deep learning in cardiac diagnosis. Across various dataset sizes, models trained on raw signal outperformed those using derived channels and minor preprocessing. The ablation analysis further revealed that derived channels have little or a slightly negative impact on model performance.

These findings highlight the capability of deep learning algorithms to extract meaningful patterns from complex physiological data without relying on hand-crafted features. This suggests a potential for a simpler, more direct approach for data input that may yield more accurate and robust models.

In future studies, we plan to further refine our approach by conducting additional experiments with 8-channel ECG data. We will focus on comparing the performance of models trained on raw 8-channel signals against those trained on extensively preprocessed 8-channel data, excluding the derived channels entirely and isolating the effects of preprocessing on the primary ECG leads.

Acknowledgements

We would like to express our gratitude to Ribeiro et al. [16] for providing the TNMG CODE dataset used in this study, which was instrumental in our analysis. This material is based upon work supported in part by the Temple University Office of the Vice President for Research, and by the Temple University College of Science and Technology Research Experience for Undergraduates program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of Temple University.

References

1. Barold, S. S. (2003). Willem Einthoven and the birth of clinical electrocardiography a hundred years ago. *Cardiac Electrophysiology Review*, *7*(1), 99–104. https://doi.org/10.1023/a:1023667812925 .

2. Bonner, R. E., Crevasse, L., Ferrer, M. I., & Greenfield, J. C. J. (1972). A new computer program for analysis of scalar electrocardiograms. *Computers and Biomedical Research, an International Journal*, *5*(6), 629–653. https://doi.org/10.1016/0010-4809(72)90043-2 .

3. Femtosim Clinical Inc. (n.d.). *Fifty Years of Physiologic Monitors*. Femtosim Clinical Inc. Retrieved May 21, 2025. url: http://www.femtosimclinical.com/History%20of%20Physiologic%20Monitors.htm .

4. Li, C., Zheng, C., & Tai, C. (1995). Detection of ECG characteristic points using wavelet transforms. *IEEE Transactions on Biomedical Engineering*, 42(1), 21–28. https://doi.org/10.1109/10.362922 .

5. Quer, G., Arnaout, R., Henne, M., & Arnaout, R. (2021). Machine Learning and the Future of Cardiovascular Care: JACC State-of-the-Art Review. *Journal of the American College of Cardiology*, *77*(3), 300–313. https://doi.org/10.1016/j.jacc.2020.11.030 .

6. Somani, S., Russak, A. J., Richter, F., Zhao, S., Vaid, A., Chaudhry, F., De Freitas, J. K., Naik, N., Miotto, R., Nadkarni, G. N., Narula, J., Argulian, E., & Glicksberg, B. S. (2021). Deep learning and the electrocardiogram: Review of the current state-of-the-art. *EP Europace*, *23*(8), 1179–1191. https://doi.org/10.1093/europace/euaa377 .

7. Kennedy, H. L. (2006). The History, Science, and Innovation of Holter Technology. *Annals of Noninvasive Electrocardiology*, *11*(1), 85–94. https://doi.org/10.1111/j.1542-474X.2006.00067.x .

8. Castells, F., Laguna, P., Sörnmo, L., Bollmann, A., & Roig, J. M. (2007). Principal Component Analysis in ECG Signal Processing. *EURASIP Journal on Advances in Signal Processing*, 2007(1), 074580. https://doi.org/10.1155/2007/74580 .

9. Abdou, A., & Krishnan, S. (2022). Horizons in Single-Lead ECG Analysis From Devices to Data. *Frontiers in Signal Processing*, *Volume 2-2022*. https://doi.org/10.3389/frsip.2022.866047.

10. Faruk, N., Abdulkarim, A., Emmanuel, I., Folawiyo, Y. Y., Adewole, K. S., Mojeed, H. A., Oloyede, A. A., Olawoyin, L. A., Sikiru, I. A., Nehemiah, M., Ya'u Gital, A., Chiroma, H., Ogunmodede, J. A., Almutairi, M., & Katibi, I. A. (2021). A comprehensive survey on low-cost ECG acquisition systems: Advances on design specifications, challenges and future direction. *Biocybernetics and Biomedical Engineering*, *41*(2), 474–502. https://doi.org/10.1016/j.bbe.2021.02.007 .

11. Berwanger, O., & Santo, K. (2022). Cardiovascular Care in Brazil: Current Status, Challenges, and Opportunities. *Circulation*, *146* (6), 435–437. https://doi.org/10.1161/CIRCULATIONAHA.122.059320 .

12. Gonçalves, M. A. A., Morais, H., Oliveira, G. M. M. de, & Mesquita, C. T. (2024). Challenges and Perspectives for Cardiology in th e Developing World: Joint Views from Africa and Latin America. *International Journal of Cardiovascular Sciences*, *37*. https://doi. org/10.36660/ijcs.20240002

13. Wagner, P., Strodthoff, N., Bousseljot, R.-D., Kreiseler, D., Lunze, F. I., Samek, W., & Schaeffter, T. (2020). PTB-XL, a large public ly available electrocardiography dataset. *Scientific Data*, *7*(1), 154. https://doi.org/10.1038/s41597-020-0495-6 .

14. Kim, Y.-G., Shin, D., Park, M. Y., Lee, S., Jeon, M. S., Yoon, D., & Park, R. W. (2017). ECG-ViEW II, a freely accessible electrocardi ogram database. *PloS One*, *12*(4), e0176222. https://doi.org/10.1371/journal.pone.0176222 .

15. Liu, H., Chen, D., Chen, D., Zhang, X., Li, H., Bian, L., Shu, M., & Wang, Y. (2022). A large-scale multi-label 12-lead electrocardio gram database with standardized diagnostic statements. *Scientific Data*, *9*(1), 272. https://doi.org/10.1038/s41597-022-01403- 5 .

16. Ribeiro, A. L. P., Paixão, G. M. M., Gomes, P. R., Ribeiro, M. H., Ribeiro, A. H., Canazart, J. A., Oliveira, D. M., Ferreira, M. P., Lima, E. M., Moraes, J. L. de, Castro, N., Ribeiro, L. B., & Macfarlane, P. W. (2019). Tele-electrocardiography and bigdata: The CODE (Clinical Outcomes in Digital Electrocardiography) study. *Journal of Electrocardiology*, 57S, S75–S78. https://doi.org/10.1016/j.jel ectrocard.2019.09.008 .

17. Ribeiro, A. H., Ribeiro, M. H., Paixão, G. M. M., Oliveira, D. M., Gomes, P. R., Canazart, J. A., Ferreira, M. P. S., Andersson, C. R., M acfarlane, P. W., Meira Jr., W., Schön, T. B., & Ribeiro, A. L. P. (2020). Automatic diagnosis of the 12-lead ECG using a deep neur al network. *Nature Communications*, 11(1), 1760. https://doi.org/10.1038/s41467-020-15432-4 .

18. *-Lead ECG Placement: The Ultimate Guide*. (n.d.). Cables and Sensors. Retrieved May 25, 2025. url: https://www.cablesandsen sors.com/pages/12-lead-ecg-placement-guide-with-illustrations#1 .

19. Hampton, J., Hampton, J., & Adlam, D. (2024). *The ECG Made Easy* (10th ed.). Elsevier. url: https://shop.elsevier.com/books/the -ecg-made-easy/hampton/978-0-323-93766-5 .

20. Burns, E., & Buttner, R. (2024, October 8). *Normal Sinus Rhythm*. Life in the Fastlane ECG Library. url: https://litfl.com/normal-si nus-rhythm-ecg-library/ .

21. Wajngarten, M., Grupi, C., Bellotti, G. M., Lemos Da Luz, P., Gastào do Serro Azul, L., & Pileggi, F. (1990). Frequency and signific ance of cardiac rhythm disturbances in healthy elderly individuals. *Journal of Electrocardiology*, *23*(2), 171–176. https://doi.org/1 0.1016/0022-0736(90)90138-R .

22. Kligfield, P., Gettes, L. S., Bailey, J. J., Childers, R., Deal, B. J., Hancock, E. W., van Herpen, G., Kors, J. A., Macfarlane, P., Mirvis, D. M., Pahlm, O., Rautaharju, P., & Wagner, G. S. (2007). Recommendations for the Standardization and Interpretation of the Ele ctrocardiogram. *Circulation*, *115*(10), 1306–1324. https://doi.org/10.1161/CIRCULATIONAHA.106.180200 .

23. Parkes, C. (2025, March 11). *EKG Interpretation, part 9: Atrioventricular Blocks*. Level Up RN. url: https://leveluprn.com/blogs/ek g-interpretation/9-atrioventricular-blocks .

24. Kusumoto, F. M., Schoenfeld, M. H., Barrett, C., Edgerton, J. R., Ellenbogen, K. A., Gold, M. R., Goldschlager, N. F., Hamilton, R. M., Joglar, J. A., Kim, R. J., Lee, R., Marine, J. E., McLeod, C. J., Oken, K. R., Patton, K. K., Pellegrini, C. N., Selzman, K. A., Thom pson, A., & Varosy, P. D. (2019). 2018 ACC/AHA/HRS Guideline on the Evaluation and Management of Patients With Bradycardia and Cardiac Conduction Delay: Executive Summary: A Report of the American College of Cardiology/American Heart Associatio n Task Force on Clinical Practice Guidelines, and the Heart Rhythm Society. *Circulation*, *140*(8), e333–e381. https://doi.org/10.1 161/CIR.0000000000000627 .

25. Burns, E., & Buttner, R. (2024, October 8). *Right Bundle Branch Block (RBBB)*. Life in the Fastlane ECG Library. url: https://litfl.c om/right-bundle-branch-block-rbbb-ecg-library/ .

26. Kashou, A.H., Shams, P. and Chhabra, L. (2024). "Right bundle branch block," in *StatPearls*, Treasure Island, FL, USA: StatPearl s Publishing. [Online]. url: https://www.ncbi.nlm.nih.gov/books/NBK507872/ .

27. Burns, E., & Buttner, R. (2024, October 8). *LeftBundle Branch Block (LBBB)*. Life in the Fastlane ECG Library. url: https://litfl.com/left-bundle-branch-block-lbbb-ecg-library/ .

28. Scherbak, D., Shams, P., & Hicks, G. (2024). *Left Bundle Branch Block (LBBB)*. StatPearls Publishing. url: https://www.ncbi.nlm.nih.gov/books/NBK482167/ .

29. Burns, E. (2024, October 8). *Sinus bradycardia*. Life in the Fastlane ECG Library. url: https://litfl.com/sinus-bradycardia-ecg-library/ .

30. Hafeez, Y., & Grossman, S. (2023). *Sinus bradycardia*. StatPearls Publishing. url: https://www.ncbi.nlm.nih.gov/books/NBK493201/ .

31. Burns, E., & Buttner, R. (2024, October 8). *Sinus Tachycardia*. Life in the Fastlane ECG Library. url: https://litfl.com/sinus-tachycardia-ecg-library/ .

32. Henning, A., & Krawiec, C. (2023). *Sinus tachycardia*. StatPearls Publishing. url: https://www.ncbi.nlm.nih.gov/books/NBK553128/ .

33. Serhal, H., Abdallah, N., Marion, J.-M., Chauvet, P., Oueidat, M., & Humeau-Heurtier, A. (2022). Overview on prediction, detection, and classification of atrial fibrillation using wavelets and AI on ECG. *Computers in Biology and Medicine*, 142, 105168. https://doi.org/10.1016/j.compbiomed.2021.105168 .

34. Hindricks, G., Potpara, T., Dagres, N., Arbelo, E., Bax, J. J., Blomström-Lundqvist, C., Boriani, G., Castella, M., Dan, G.-A., Dilaveris, P. E., Fauchier, L., Filippatos, G., Kalman, J. M., La Meir, M., Lane, D. A., Lebeau, J.-P., Lettino, M., Lip, G. Y. H., Pinto, F. J., … Watkins, C. L. (2021). 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *European Heart Journal*, 42(5), 373–498. https://doi.org/10.1093/eurheartj/ehaa612.

35. Joglar, J. A., Chung, M. K., Armbruster, A. L., Benjamin, E. J., Chyou, J. Y., Cronin, E. M., Deswal, A., Eckhardt, L. L., Goldberger, Z. D., Gopinathannair, R., Gorenek, B., Hess, P. L., Hlatky, M., Hogan, G., Ibeh, C., Indik, J. H., Kido, K., Kusumoto, F., Link, M. S., … Van Wagoner, D. R. (2024). 2023 ACC/AHA/ACCP/HRS Guideline for the Diagnosis and Management of Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation*, 149(1), e1–e156. url: https://www.ahajournals.org/doi/ https://doi.org/10.1161/CIR.0000000000001193 .

36. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90 .

37. Thundiyil, S. C., & Picone, J. (2025). Time Series Analysis from Classical Methods to Transformer-Based Approaches: A Review. In *Signal Processing in Medicine and Biology: Applications of Artificial Intelligence in Medicine and Biology* (Vol. 1, p. 56). Springer. url: https://isip.piconepress.com/publications/book_sections/2024/springer/transformers/ . (in publication).

38. Pastika, L., Sau, A., Patlatzoglou, K., Sieliwonczyk, E., Ribeiro, A. H., McGurk, K. A., Khan, S., Mandic, D., Scott, W. R., Ware, J. S., Peters, N. S., Ribeiro, A. L. P., Kramer, D. B., Waks, J. W., & Ng, F. S. (2024). Artificial intelligence-enhanced electrocardiography derived body mass index as a predictor of future cardiometabolic disease. *Npj Digital Medicine*, 7(1), 167. https://doi.org/10.1038/s41746-024-01170-0 .

39. Sau, A., Sieliwonczyk, E., Patlatzoglou, K., Pastika, L., McGurk, K. A., Ribeiro, A. H., Ribeiro, A. L. P., Ho, J. E., Peters, N. S., Ware, J. S., Tayal, U., Kramer, D. B., Waks, J. W., & Ng, F. S. (2025). Artificial intelligence-enhanced electrocardiography for the identification of a sex-related cardiovascular risk continuum: A retrospective cohort study. *The Lancet Digital Health*, 7(3), e184–e194. https://doi.org/10.1016/j.landig.2024.12.003 .

40. von Bachmann, P., Gedon, D., Gustafsson, F. K., Ribeiro, A. H., Lampa, E., Gustafsson, S., Sundström, J., & Schön, T. B. (2024). Evaluating regression and probabilistic methods for ECG-based electrolyte prediction. *Scientific Reports*, 14(1), 15273. https://doi.org/10.1038/s41598-024-65223-w .

41. Kung, B.-H., Hu, P.-Y., Huang, C.-C., Lee, C.-C., Yao, C.-Y., & Kuan, C.-H. (2021). An Efficient ECG Classification System Using Resource-Saving Architecture and Random Forest. *IEEE Journal of Biomedical and Health Informatics*, 25(6), 1904–1914. https://doi.org/10.1109/JBHI.2020.3035191 .

42. Melgani, F., & Bazi, Y. (2008). Classification of Electrocardiogram Signals With Support Vector Machines and Particle Swarm Optimization. *IEEE Transactions on Information Technology in Biomedicine*, 12(5), 667–677. https://doi.org/10.1109/TITB.2008.923147 .

43. Turnip, A., Ilham Rizqywan, M., Kusumandari, D. E., Turnip, M., & Sihombing, P. (2018). Classification of ECG signal with Support Vector Machine Method for Arrhythmia Detection. *Journal of Physics: Conference Series*, 970(1), 012012. https://doi.org/10.1088/1742-6596/970/1/012012.

44. Reddy, S. D., Murugan, R., Nandi, A., & Goel, T. (2023). Classification of arrhythmia disease through electrocardiogram signals using sampling vector random forest classifier. *Multimedia Tools and Applications*, 82(17), 26797–26827. https://doi.org/10.1007/s11042-022-14304-x

45. Y. Zhang, J. Li, S. Wei, F. Zhou, & D. Li. (2021). Heartbeats Classification Using Hybrid Time-Frequency Analysis and Transfer Learning Based on ResNet. *IEEE Journal of Biomedical and Health Informatics*, 25(11), 4175–4184. https://doi.org/10.1109/JBHI.2021.3085318 .

46. Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adam, M., Gertych, A., & Tan, R. S. (2017). A deep convolutional neural network model to classify heartbeats. *Computers in Biology and Medicine*, 89, 389–396. https://doi.org/10.1016/j.compbiomed.2017.08.022

47. Kiranyaz, S., Ince, T., & Gabbouj, M. (2016). Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Transactions on Biomedical Engineering*, 63(3), 664–675. https://doi.org/10.1109/TBME.2015.2468589 .

48. Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25(1), 65–69. https://doi.org/10.1038/s41591-018-0268-3 .

49. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 .

50. S. Saadatnejad, M. Oveisi, & M. Hashemi. (2020). LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices. *IEEE Journal of Biomedical and Health Informatics*, *24*(2), 515–523. https://doi.org/10.1109/JBHI.2019.2911367 .

51. Cheng, J., Zou, Q., & Zhao, Y. (2021). ECG signal classification based on deep CNN and BiLSTM. *BMC Medical Informatics and Decision Making*, 21(1), 365. https://doi.org/10.1186/s12911-021-01736-y .

52. J. Hu, W. L. Goh, & Y. Gao. (2024). LSTM-based ECG Signal Classification with Multi-level One-hot Encoding for Wearable Applications. *2024 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 1–5. https://doi.org/10.1109/BioCAS61083.2024.10798273 .

53. Gagliano, L., Bou Assi, E., Nguyen, D. K., & Sawan, M. (2019). Bispectrum and Recurrent Neural Networks: Improved Classification of Interictal and Preictal States. *Scientific Reports*, 9(1), 15649. https://doi.org/10.1038/s41598-019-52152-2.

54. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, \Lukasz, & Polosukhin, I. (2017). Attention is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–6010. url: https://dl.acm.org/doi/ https://doi.org/10.5555/3295222.3295349 .

55. Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023). Are Transformers Effective for Time Series Forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(9), 11121–11128. https://doi.org/10.1609/aaai.v37i9.26317 .

56. Cao, H., Huang, Z., Yao, T., Wang, J., He, H., & Wang, Y. (2023). InParformer: Evolutionary Decomposition Transformers with Interactive Parallel Attention for Long-Term Time Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(6), 6906–6915. https://doi.org/10.1609/aaai.v37i6.25845 .

57. Weimann, K., & Conrad, T. O. F. (2021). Transfer learning for ECG classification. *Scientific Reports*, 11(1), 5251. https://doi.org/10.1038/s41598-021-84374-8 .

58. Van Steenkiste, G., van Loon, G., & Crevecoeur, G. (2020). Transfer Learning in ECG Classification from Human to Horse Using a Novel Parallel Neural Network Architecture. *Scientific Reports*, 10(1), 186. https://doi.org/10.1038/s41598-019-57025-2 .

59. Ben-Moshe, N., Tsutsui, K., Brimer, S. B., Zvuloni, E., Sörnmo, L., & Behar, J. A. (2024). RawECGNet: Deep Learning Generalization for Atrial Fibrillation Detection From the Raw ECG. *IEEE Journal of Biomedical and Health Informatics*, *28*(9), 5180–5188. https://doi.org/10.1109/JBHI.2024.3404877 .

60. Xu, S. S., Mak, M.-W., & Cheung, C.-C. (2019). Towards End-to-End ECG Classification With Raw Signal Extraction and Deep Neural Networks. *IEEE Journal of Biomedical and Health Informatics*, *23*(4), 1574–1584. https://doi.org/10.1109/JBHI.2018.2871510 .

61. Moons, K. G. M., Damen, J. A. A., Kaul, T., Hooft, L., Andaur Navarro, C., Dhiman, P., Beam, A. L., Van Calster, B., Celi, L. A., Denaxas, S., Denniston, A. K., Ghassemi, M., Heinze, G., Kengne, A. P., Maier-Hein, L., Liu, X., Logullo, P., McCradden, M. D., Liu, N., … van Smeden, M. (2025). PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ*, *388*. https://doi.org/10.1136/bmj-2024-082505.

62. Salahuddin, Z., Woodruff, H. C., Chatterjee, A., & Lambin, P. (2022). Transparency of deep neural networks for medical image analysis: A review of interpretability methods. *Computers in Biology and Medicine*, *140*, 105111. https://doi.org/10.1016/j.compbiomed.2021.105111 .

63. Purba, S. A., Melles, A.-M., Hackel, D., Obeid, I., & Picone, J. (2025). Assessing the Visual Reasoning of Multimodal Language Models in Biomedical Applications. *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, *1*, 1–5. url: www.isip.piconepress.com/publications/unpublished/conferences/2025/ieee_spmb/segmentation/ . (in review).

64. Singh, H. (2025, May 31). *10 Techniques to Solve Imbalanced Classes in Machine Learning (Updated 2025)*. Analytics Vidhya. url: https://www.analyticsvidhya.com/blog/2020/07/10-techniques-to-deal-with-class-imbalance-in-machine-learning/ .

65. Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2010). RUSBoost: A Hybrid Approach to Alleviating Class Imbalance. *IEEE Transactions on Systems, Man and Cybernetics. Part A, Systems and Humans, 2010, Vol.40 (1), p.185–197*. https://doi.org/10.1109/TSMCA.2009.2029559 .

66. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1109/TSMCA.2009.2029559 .

67. Kingma, D. P., & Ba, J. L. (2015). Adam: A Method for Stochastic Optimization. *Proceedings of the International Conference on Learning Representations*, 1–15. url: https://arxiv.org/abs/1412.6980 .

68. Mao, A., Mohri, M., & Zhong, Y. (2023). Cross-Entropy Loss Functions: Theoretical Analysis and Applications. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, & J. Scarlett (Eds.), *Proceedings of the 40th International Conference on Machine Learning* (Vol. 202, pp. 23803–23828). PMLR. url: https://proceedings.mlr.press/v202/mao23b.html .

69. Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. url: http://www.amazon.com/Introduction-Information-Retrieval-Christopher-Manning/dp/0521865719 .

70. Meyes, R., Lu, M., Puiseau, C. W. de, & Meisen, T. (2019). *Ablation Studies in Artificial Neural Networks*. url: https://arxiv.org/abs/1901.08644 .

71. Graham, M. H. (2003). Confronting Multicollinearity in Ecological Multiple Regression. *Ecology*, *84*(11), 2809–2815. https://doi.org/10.1890/02-3114 .

72. Belsley, D. A. (n.d.). *Regression diagnostics: Identifying influential data and sources of collinearity/David A. Belsley, Edwin Kuh, Roy E. Welsch*. New York : Wiley, c2004. https://doi.org/10.1002/0471725153 .

© Springer Nature