POLITECNICO DI TORINO

Master of Science Degree in Computer Engineering

Master's Degree Thesis

Automated Detection of ECG Electrode Interchange Using Machine Learning



Supervisors
Prof. Vincenzo Randazzo
Prof. Eros Gian Alessandro Pasero
Supervisors' signatures

. . . . . .
. . . . . .

Candidates
Carmen Sonia Altieri
Gaetano Caruso
Candidates' signatures

. . . . . .
. . . . . .

Academic Year 2025–2026

To my father:
in every place
I wander,
in every pieces
of this world,
I will be looking
for traces of you.

Acknowledgements

My first thank you goes to Professors Randazzo and Pasero, who gave me the opportunity to work on this thesis. I would like to express my gratitude for their guidance throughout this work.

A special thank you goes to Dr.Casella who constantly supported me, with patience and valuable advice, throughout this journey.

I would also like to thank Aizoi and Eng. Manzini for providing the data and the experience that made this work possible.

I would also like to thank my family, especially my parents, who always supported and encouraged me. They believed in myself even when I didn't.

To my sister Claudia, thank you for always being there for me, even when I wasn't very easy to be with. Thank you for sharing both the good moments and the difficult ones.

Finally, I would like to thank my friends, both the ones I have always known and the ones I shared university with. Thank you all for the laughter, the conversations and the shared struggles.

This thesis is the conclusion of an important chapter of my life, but it is also the result of all the people and the experiences that accompanied me along the way. For this, I am truly grateful.

Contents

List of Tables 8

List of Figures 10

1 Introduction 3
- 1.1 Clinical Context and ECG Acquisition . . . 4
- 1.2 Errors During ECG Acquisition . . . 5
- 1.3 Thesis Organization . . . 6

2 Background and Related Work 7
- 2.1 Electrocardiographic Background . . . 7
  - 2.1.1 Electrical Origin and Morphology of the ECG Signal . . . 7
  - 2.1.2 The Standard 12-Lead ECG System . . . 9
  - 2.1.3 Cardiac Electrical Axis and Inter-Subject Variability . . . 12
  - 2.1.4 Standard Electrode Placement . . . 14
- 2.2 ECG Acquisition Errors . . . 16
  - 2.2.1 Limb Electrode Interchange . . . 16
  - 2.2.2 Precordial Electrode Interchange . . . 22
- 2.3 Automatic Detection of Electrode Errors . . . 22
  - 2.3.1 Rule-Based and Inter-Lead Correlation Methods . . . 23
  - 2.3.2 Feature Engineering and Traditional Machine Learning . . . 24
  - 2.3.3 Deep Learning Approaches . . . 24
- 2.4 Data Simulation and Signal Quality in Previous Studies . . . 25
  - 2.4.1 Simulation of Electrode Interchanges . . . 25
  - 2.4.2 Signal Quality Assessment and Preprocessing . . . 26
- 2.5 Research Gaps and Positioning of the Thesis . . . 27

3 Methods and Proposed Solution 29
- 3.1 Signal Preprocessing Pipeline . . . 29
  - 3.1.1 Raw Signal Structure, Quality Control and Label Normalization . . . 30

3.1.2 Power-Line Removal, Digital Filtering and Polyphase Resampling . . . 31
3.1.3 Signal Quality Assessment (SQA) and Structural Detection of the Neutral Electrode . . . 36
3.1.4 Normalization via Robust Global Z-Score . . . 38
3.2 Dataset and Sampling Strategy . . . 40
3.2.1 Initial ECG Dataset and Report Organization . . . 40
3.2.2 Collection, Selection, and Integrity Filters of the Initial Clinical Data . . . 43
3.2.3 Common Preprocessing Pipeline, Quality Assessment, and HDF5 Construction . . . 44
3.2.4 Dataset Creation from the Common Pipeline . . . 45
3.2.5 Label Encoding . . . 46
3.2.6 Creation of the Baseline Dataset . . . 47
3.2.7 Real Clinical Test Set for Sim-to-Real Validation . . . 48
3.2.8 Pathology Dataset for Axis Robustness Stress-Test . . . 49
3.2.9 Dataset for Precordial Leads . . . 50
3.3 Neural Network Architectures and Training . . . 51
3.3.1 LDenseNet Architecture for Limb-Lead Inversion Detection . 52
3.3.2 Inter Lead Correlation Model for Precordial Leads Interchange Detection . . . 54
3.3.3 Training Procedure . . . 57
3.3.4 Evaluation Protocol . . . 62
3.4 The Geometric Limit of Limb Leads and V6 . . . 63
3.4.1 Global Table of Angular Transformations . . . 64
3.4.2 Vector Mimicry and Information Collapse . . . 65
3.4.3 Mathematical Model of the Transformations . . . 65
3.4.4 General Law of Mimicry . . . 66
3.4.5 Eigen-Space and Null Space . . . 66
3.4.6 Analytical Interpretation of the Role of V6 . . . 69
3.4.7 The Absolute Limit: when V6 is not enough . . . 74
3.4.8 Clinical and Theoretical Interpretation . . . 74

4 Experimental Results 77
4.1 Evaluation Metrics . . . 77
4.1.1 Limb Leads: Simulated-to-Simulated Scenario . . . 79
4.1.2 Effect of Semi-Supervised Training . . . 80
4.1.3 Precordial Leads: Simulated-to-Simulated Scenario . . . 82
4.2 Simulated-to-Real Scenario: 6-lead and 7-lead Models . . . 84
4.2.1 Lightweight Complexity Reduction . . . 87
4.3 Summary of Results and Findings . . . 89

5 General Discussion 91
5.1 Performance in the Simulated-to-Simulated Setting 92
5.2 Effect of ECG-level Probability Aggregation 93
5.3 Contribution of the Semi-Supervised Extension 93
5.4 Generalization from Simulated to Real Clinical ECGs 94
5.5 Geometric Interpretation of the Contribution of V6 95
5.6 Error Interpretation and Uncertainty of the Clinical Ground Truth . 95
5.7 Clinical Interpretation of the Binary Task 96
5.8 Lightweight Input Configurations 96
5.9 Methodological Strengths and Limitations 97
5.10 Overall Interpretation 97

6 Conclusions and Future Work 99
6.1 Conclusions 99
6.2 Future Work 100

List of Tables

3.1 Invalidation criteria for time windows at Level 2 of the Signal Quality Assessment. . . . 37
3.2 Class encoding and synthetic transformations adopted for the limb-lead classification task. The prime symbol denotes the transformed lead. . . . 46
3.3 Class encoding adopted for the precordial-lead classification task. . . . 47
3.4 Distribution of categories in the real limb-lead dataset in order of occurrence. . . . 48
3.5 Main architectural characteristics of the implemented models. . . . 56
3.6 Training configurations adopted for the three experiments. . . . 61
3.7 Transformation of standard cardiac axes for the different limb-lead inversion classes. . . . 64
3.8 Extension of angular transformations to the axis $\alpha = +120°$. . . . 64
3.9 Nine points of perfect collapse (Null Space) between pairs of limb-lead transformations. . . . 70
4.1 Per-class performance on limb leads (simulated-to-simulated scenario, window level). . . . 79
4.2 Per-class performance on limb leads (simulated-to-simulated scenario, patient level). . . . 80
4.3 Aggregate performance on limbs simulated-to-simulated (window level): baseline vs SSL model. . . . 81
4.4 Aggregate performance on limbs simulated-to-simulated (patient level): baseline vs SSL model. . . . 82
4.5 Per-class performance on precordial leads (simulated-to-simulated scenario, window level). . . . 83
4.6 Per-class performance on precordial leads (simulated-to-simulated scenario, patient level). . . . 84
4.7 Aggregate performance in the simulated-to-real scenario at ECG level for the 6-lead and 7-lead configurations. . . . 86

4.8 Class-wise errors in the simulated-to-real evaluation. Change is computed as the number of 6-lead errors minus the number of 7-lead errors; positive values indicate fewer errors with the 7-lead configuration. 86
4.9 Per-class metrics in the simulated-to-real scenario (patient level). . . 87
4.10 Patient-level accuracy for different lead configurations. 89

List of Figures

1.1 CardioPAD PRO device used for ECG acquisition in a telemedicine setting AIZOI [2026]. 4
2.1 Main waves, segments, and intervals of a standard ECG recording Wikipedia contributors [2026]. 8
2.2 Main steps of cardiac cycle Kuhn and Rose [2008]. 8
2.3 Combined representation of the frontal and horizontal reference planes of the standard 12-lead ECG Kania et al. [2020]. 10
2.4 Einthoven's triangle and the standard limb-lead axes in the hexaxial reference system Wikimedia Commons [2026]. 11
2.5 Schematic representation of the cardiac electrical axis in the hexaxial reference system Kuhn and Rose [2008]. 13
2.6 Standard anatomical placement of the six precordial electrodes Dreamstime [2026]. 15
2.7 Effects of selected limb-electrode interchanges not involving the RL electrode Batchvarov et al. [2007]. 17
2.8 Effects of selected limb-electrode interchanges involving the RL electrode Batchvarov et al. [2007]. 18
3.1 Example of *raw* 30-second ECG printout with the 12 leads ordered according to the convention adopted in the pipeline. 32
3.2 Effect of the 50 Hz notch filter on power-line interference. 34
3.3 Effect of the band-pass filter (0.5–120 Hz) on baseline drift. 34
3.4 Effect of polyphase resampling: comparison between the original signal (1000 Hz) and the resampled signal (250 Hz). 35
3.5 LDenseNet architecture adopted for limb-lead inversion detection. The architecture was adapted from Huang et al. Huang et al. [2024]. 53
3.6 ILC architecture adopted for precordial-lead inversion detection. Each precordial lead is processed by an LDenseNet branch with shared weights. The extracted feature maps are combined with the inter-lead correlation coefficients and passed to the final classifier. The architecture was adapted from Huang et al. Huang et al. [2024]. 54
3.7 Family 2 under the General Law (LA–RA vs ROT_ANT). 67

3.8 Family 1 under the General Law: polar diagram and apparent frontal tracings. 68
3.9 Family 2 in Eigen-space: polar representation and apparent frontal tracings. 71
3.10 Family 1 in Eigen-space: polar diagram and apparent frontal tracings. 72
3.11 Family 3 in Eigen-space: polar diagram and apparent frontal tracings. 73
4.1 Confusion matrices for limb leads, simulated-to-simulated scenario. 81
4.2 Confusion matrices for limb leads, SSL scenario. 83
4.3 Confusion matrices for precordial leads. 85
4.4 Confusion matrices in the simulated-to-real scenario for the 6-lead and 7-lead models, trained on the pathology dataset. 88

Abstract

This thesis investigates the development of a deep learning pipeline for the detection of electrode interchange in 12-lead electrocardiograms (ECGs). Recognizing the incorrect placement of the electrodes is vital, since electrode interchange can significantly alter signal morphology, potentially mimicking or concealing clinically relevant cardiac conditions.

Limb and precordial leads were treated as two separate problems with corresponding separate pipelines. At first, the ECG undergoes a signal preprocessing pipeline, which includes power-line interference removal, band-pass filtering, resampling, signal quality assessment, robust normalization, and segmentation into two-second windows. Limb electrode interchange was formulated as a six class classification task, including the normal configuration and five interchange configurations. For the precordial interchange, sixteen classes were used, including the normal configuration and fifteen pairwise exchanges among leads V1-V6. Two different architectures were used: a lightweight model called LDenseNet for the limb task and the Inter Lead Correlation model for the precordial task, which combined single lead feature extraction with the correlation between leads. Then, a semi-supervised extension based on pseudo labelled clinical recordings was also employed.

Results show that on simulated data, the models achieved ECG-level multiclass accuracies of 97.82% for the limb-lead interchange task and 98.49% for the precordial interchange task. The semi-supervised approach provided a slight improvement over supervised limb training. The simulated to real task achieved an accuracy of 89.28% using the six limb leads and then reached 90.44% when V6 was included. These results underline the domain shift between simulated and real data and indicate the contribution of V6.

These results highlight the potential of deep learning models in the automatic electrode interchange detection and quality check tool. After further validation, such models could be integrated into ECG acquisition systems to provide real time feedback and alert the operators in case of suspected electrode interchanges, so that only correct acquired ECGs are transmitted for the clinical interpretation. This could reduce the need of repeated recordings and limit the risk of diagnostic errors caused by electrode interchange.

Chapter 1

Introduction

In recent years, technological development has increased the availability of medical devices capable of remotely acquiring ECG records and transmitting them to a specialist for evaluation. This made ECG acquisition possible from different places outside conventional hospitals, like pharmacies and patients' homes. In this context, the reliability of the signal, strongly relies on a correct acquisition. Indeed, incorrect electrode connection or placement may alter waveform morphology, polarity, and inter-lead relationships, potentially reducing the diagnostic value of the ECG and requiring the examination to be repeated Rjoob et al. [2020], Huang et al. [2024].

The objective of this thesis is to explore the automatic detection of electrode interchanges in standard 12-lead ECG recordings. This work relies on two separate multi-class classification tasks for the limb-electrode and precordial-electrode interchanges, since these produce different electrical and morphological effects. Anatomical precordial misplacement is discussed as a separate category of acquisition errors, but it is not the target analysis of the present work, particularly due to the scarcity of the data.

This analysis was conducted with the collaboration of Aizoi, a Turin-based company founded in 2011 as the evolution of Biotechware, which develops integrated telemedicine solutions for cardiovascular diagnostics. An important contribution was given by the company's CTO and co-founder Ing. Enrico Manzini.

The proposed approach combines ECG signal preprocessing with a deep learning architecture for the two classification tasks, in order to detect the particular interchange class. Furthermore, particular attention is given to the broader distinction between normal and abnormal acquisitions, because the main practical requirement is to prevent a wrongly acquired ECG from being accepted as correct. The final aim is to propose this work as a unified preliminary pipeline, which is able to process the ECG record in a telemedicine workflow, without replacing the clinical interpretation performed by the specialist.

1.1 Clinical Context and ECG Acquisition

The electrocardiogram (ECG) is a diagnostic examination that records the electrical activity of the heart through electrodes placed on the surface of the body. It captures the temporal and spatial variations of cardiac electrical potentials, and due to its accessibility, low cost and its diagnostic value, the ECG is one of the most widely used tools for the assessment of the cardiac function De Coster et al. [2025].

The mechanism of the electrocardiogram (ECG) was discovered by William Einthoven, who introduced the string galvanometer at the beginning of the twentieth century. Since then, technological progress has led to compact digital devices that can acquire, store, process and transmit ECG signal in a short period of time De Coster et al. [2025].

A standard 12-lead ECG is obtained by placing four limb electrodes and six precordial electrodes on the surface of the body. The potentials measured by these electrodes create six limb leads, that describe the heart electrical activity in the frontal plane, and six precordial leads, which describe the same activity but in the horizontal plane. Since the information provided by the ECG is derived from the electrodes, their placement is fundamental for a correct signal acquisition.

An example of this evolution in telemedicine is the CardioPAD PRO, a portable ECG device developed by Aizoi and fully integrated in the Biotecware cloud-ecosystem.



Figure 1.1: CardioPAD PRO device used for ECG acquisition in a telemedicine setting AIZOI [2026].

Remote ECG acquisition improves accessibility and may reduce the time required to obtain a medical report. However, it may also increase the risk of acquisition errors, since ECG recordings can be performed by operators with different levels of expertise. Although electrode interchange is a relatively infrequent event,

with a reported incidence ranging from 0.4% to 4%, the large number of ECGs acquired in clinical practice means that a considerable number of recordings may still be affected Huang et al. [2024]. The incidence is even higher in non-clinical or less controlled acquisition settings. For this reason, acquisition errors should be detected as early as possible before the ECG is transmitted for clinical evaluation.

1.2 Errors During ECG Acquisition

ECG signal quality and therefore, its reliability, depends on several factors, such as patient preparation, electrode contact, cable connection, signal quality, and anatomical placement. Errors that occur during this stage may alter the signal morphology, amplitude or polarity, and consequently its clinical relevance.

There are different types of acquisition errors. A first category regards errors that affect the signal quality, for example, motion artifacts, poor electrode-to-skin contact, baseline wander, muscle interference, and electrical noise. These disturbances may obscure relevant ECG components and therefore, the signal could not be analysed.

Then there are problems regarding the placing of the electrodes on the skin: these can either be interchanged or misplaced. Electrode interchange occurs when the operator performing the ECG, interchanges two electrodes, and although the electrical potentials measured are correct, they are associated to wrong leads in the recorded ECG. Electrode misplacement occurs when the electrodes are positioned away from the correct anatomical location, so the signal produced is acquired from a non standard location. Even with a millimetric offset, the recorded ECG may result altered, since the signal is acquired from a non standard position.

If these errors regard limb leads, there could be problems affecting more leads simultaneously, producing changes in waveform polarity, electrical axis, and inter-lead relationships. An interchange between precordial leads may disrupt the natural spatial progression from V1 to V6.

These alterations are clinically relevant because ECG, that are wrongly acquired, could resemble true cardiac abnormalities, such as ventricular hypertrophy, intraventricular conduction disorders, myocardial ischemia, or myocardial infarction. In these cases, distinguishing a technical error from a true pathological pattern may be particularly difficult even for an expert cardiologist Burns and Buttner [2024].

The consequences could include the repetition of the ECG, delays in clinical assessment, unnecessary diagnostic procedures, and, in the most critical cases, inappropriate clinical decisions. For this reason, ECG technical correctness should be assessed before the recording is transmitted to a specialist for evaluation.

Several manual and automatic criteria have been proposed to identify these

problems. Traditional approaches focus more on the morphology of the signal, such as expected waveform polarities, amplitude relationships and flat leads. Later, machine learning methods were proposed, using classifiers based on manually extracted statistical and morphological features. More recently, deep learning approaches have demonstrated the ability to learn discriminative representations directly from ECG signals and to distinguish correct acquisitions from different interchange configurations Rjoob et al. [2020, 2021], Huang et al. [2024].

1.3 Thesis Organization

The remainder of this thesis is organized into five more chapters:

- Chapter 2 presents a theoretical panoramic on the ECG, required to understand the effect of electrode interchanges. It also reviews the state of the art methods, including traditional, but also machine learning and deep learning approaches to deal with the electrode-interchange problems.
- Chapter 3 describes the proposed preprocessing and classification pipeline. It introduces the datasets, signal preprocessing, window generation, simulation of electrode-interchange configurations, neural architectures, training procedure, and evaluation protocol together with a mathematical/geometrical analysis of the errors.
- Chapter 4 reports the results obtained for the limb-electrode and precordial-electrode interchange tasks. Performance is analysed through a multi-class classification and the broader normal-versus-abnormal in both windowed and patient-whole analysis. Furthermore, results on geometric observations about the electrical heart axis and its inclinations are shown.
- Chapter 5 presents a discussion and an interpretation of the main experimental findings. This chapter compares and analyses the results obtained for the limb and the precordial tasks. It also discusses the generalization of the proposed models from simulated to real ECG recordings, the contribution of the additional precordial lead, and the main limitations of the proposed approach.
- Finally, Chapter 6 summarizes the main contributions and findings of the thesis, also discussing the possible directions for future work.

Chapter 2

Background and Related Work

This chapter presents the related work and the theoretical background useful for understanding the electrode-interchange problem addressed in this thesis. The first part describes the electrical origin and morphology of the ECG signal, the standard 12-lead system, and the anatomical placement of the limb and precordial electrodes. Then, the main acquisition problems are described, along with their effects on the ECG record. Finally, the state of the art is introduced, from manual clinical rules to deep learning approaches.

2.1 Electrocardiographic Background

2.1.1 Electrical Origin and Morphology of the ECG Signal

An ECG measures the spatial variations in electrical potential generated by the heart. This is caused by the depolarization and repolarization of atrial and ventricular myocardial cells. De Coster et al. [2025].

Each ECG lead measures the projection of the instantaneous cardiac electrical vector along a specific observation axis. The same cardiac event produces different amplitudes, polarities and morphologies depending on the lead from which it is observed. That is why the information about the electrical field generated by the heart cannot be obtained from a single lead, but only with the simultaneous analysis of multiple leads.

The main components of the ECG are the P wave, the PR interval, the QRS complex, the ST segment, the T wave, and the QT interval. Each of these components is associated with a different phase of cardiac electrical activation or recovery Surawicz and Knilans [2008].

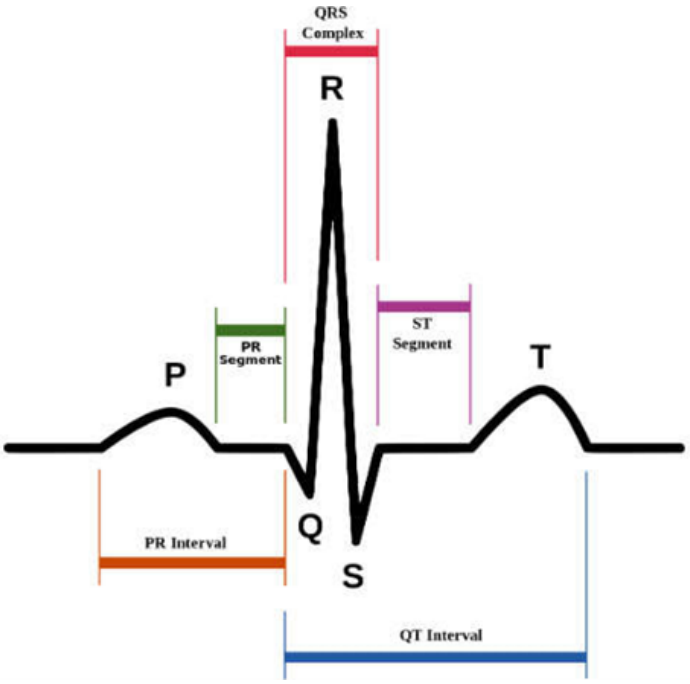


Figure 2.1: Main waves, segments, and intervals of a standard ECG recording Wikipedia contributors [2026].

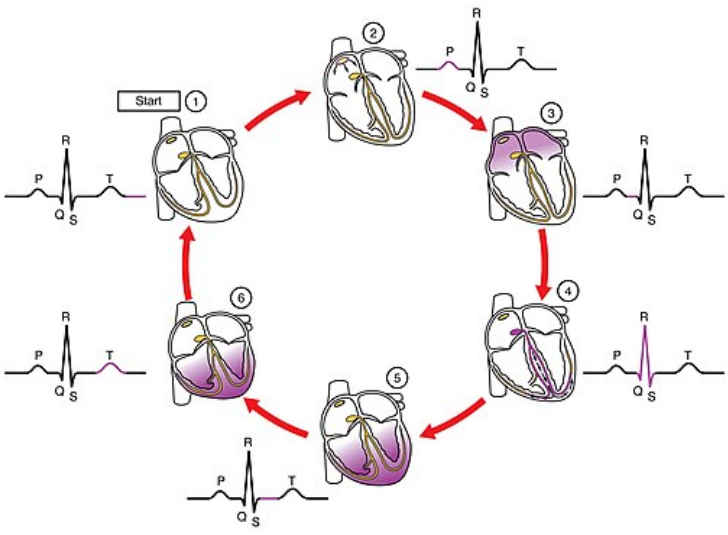


Figure 2.2: Main steps of cardiac cycle Kuhn and Rose [2008].

Figure 2.1 illustrates the main clinically relevant components of a standard ECG waveform and their relationship to the phases of the cardiac cycle shown in Figure 2.2. The following paragraph briefly describe each ECG component.

The P wave comes from atrial depolarization. It generally presents a smooth and rounded morphology and it is commonly positive in leads I and II. Its amplitude and polarity depend on the lead orientation and on the direction of atrial activation. The PR interval extends from the P wave to the QRS complex and includes conduction through the atria, the atrioventricular node, and the specialized ventricular conduction system. At normal sinus rates, its duration is generally between 120 and 200 ms Hancock et al. [2009], Surawicz and Knilans [2008].

The QRS complex represents ventricular depolarization and since the left ventricle is greater that the right ventricle, the mean ventricular electrical vector is generally directed downwards and to the left. Under normal conditions, the QRS complex has a duration shorter than 120 ms, although its amplitude, polarity, and morphology vary considerably among the different leads Surawicz et al. [2009]. In the precordial leads, ventricular depolarization produces a spatial progression from V1 to V6. The right precordial leads usually have a dominant S wave, while the R wave amplitude progressively increases towards the left precordial leads. Between V_2 and V_4, there is a transition zone in which the R wave becomes larger than the S wave Macfarlane et al. [2010].

The QRS complex is followed by the ST segment, which represents the interval during which most of the ventricular myocardium is depolarised and it is normally close to the isoelectric baseline. The T wave corresponds to ventricular repolarization and it generally follows the dominant direction of the QRS complex in most leads Vink et al. [2018], Pueyo et al. [2016].

These morphological components are particularly relevant to understand the electrode-interchange error. When an electrode is exchanged with another, the direction from which the cardiac electrical activity is observed changes, and therefore, there could be changes in polarity, amplitude and morphology of the lead.

2.1.2 The Standard 12-Lead ECG System

A standard 12-lead ECG is obtained placing ten physical electrodes on the body surface: four limb electrodes and six precordial electrodes. The limb electrodes are identified as right arm (RA), left arm (LA), right leg (RL), and left leg (LL), whereas the six precordial electrodes are identified as V_1 to V_6.

The four limb electrodes are used to obtain six limb leads, while the six precordial electrodes contribute to six precordial leads. The electrode is the physical sensor that measures the electrical potential, while the lead is the signal constructed from one or more electrode potentials.

The six limb leads describe the heart electrical activity on the frontal plane,

while the precordial leads provide complementary information in the horizontal plane. All together they provide a multidirectional representation of the same cardiac electrical information:

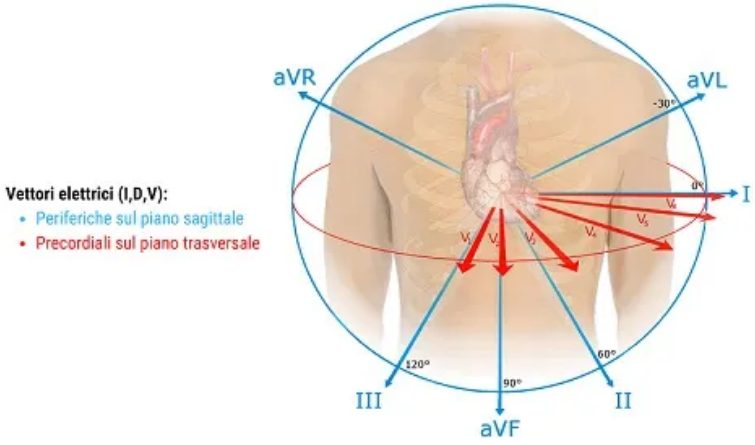


Figure 2.3: Combined representation of the frontal and horizontal reference planes of the standard 12-lead ECG Kania et al. [2020].

Bipolar Limb Leads and Einthoven's Triangle

The limb leads I, II and III are defined as potential differences between pairs of limb electrodes:

$$I = V_{\mathrm{LA}} - V_{\mathrm{RA}}, \tag{2.1}$$
$$II = V_{\mathrm{LL}} - V_{\mathrm{RA}}, \tag{2.2}$$
$$III = V_{\mathrm{LL}} - V_{\mathrm{LA}}. \tag{2.3}$$

Geometrically, these relationships can be represented through Einthoven's triangle, whose vertices correspond to RA, LA and LL. Lead I is directed from the right arm towards the left arm, lead II from the right arm towards the left leg, and lead III from the left arm towards the left leg Barold [2010].

Einthoven's law states that:

$$II = I + III. \tag{2.4}$$

This equation surely expresses the algebraic redundancy of the limb-lead system, but it cannot be used alone to detect electrode interchange. This is because the equation follows from the definition of the leads, so it could remain valid even though the electrodes are not where they are supposed to be. The equation should

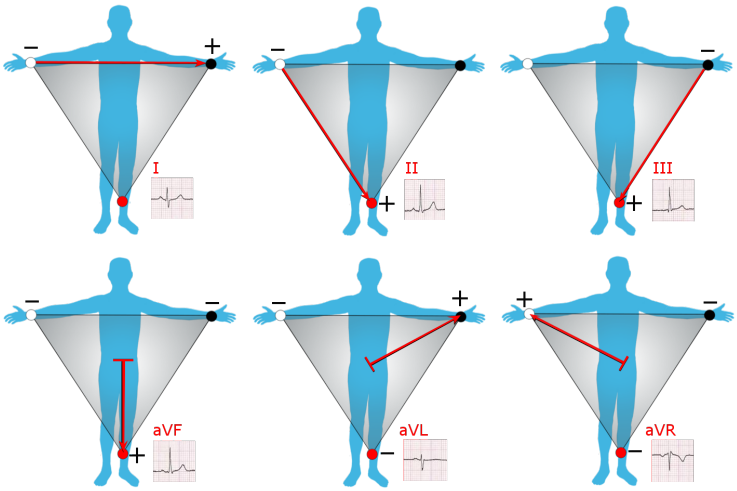


Figure 2.4: Einthoven's triangle and the standard limb-lead axes in the hexaxial reference system Wikimedia Commons [2026].

be used together with analysis on the waveform polarity and morphology, and all leads should be considered.

Other frontal plane observation directions are provided by the augmented limb leads aVR, aVL, aVF. They measure the potential of a limb electrode with respect to the average potential of the other two:

$$aVR = V_{\mathrm{RA}} - \frac{V_{\mathrm{LA}} + V_{\mathrm{LL}}}{2}, \tag{2.5}$$

$$aVL = V_{\mathrm{LA}} - \frac{V_{\mathrm{RA}} + V_{\mathrm{LL}}}{2}, \tag{2.6}$$

$$aVF = V_{\mathrm{LL}} - \frac{V_{\mathrm{RA}} + V_{\mathrm{LA}}}{2}. \tag{2.7}$$

Precordial Leads and the Wilson Central Terminal

The six precordial leads provide a representation of the heart electrical activity in the horizontal plane. Each precordial electrode measures the potential difference with respect to a common reference, called Wilson central terminal (WCT):

$$V_{\mathrm{WCT}} = \frac{V_{\mathrm{RA}} + V_{\mathrm{LA}} + V_{\mathrm{LL}}}{3}. \tag{2.8}$$

Consequently, each precordial lead can be expressed as:

$$V_i = V_{\mathrm{electrode},i} - V_{\mathrm{WCT}}, \qquad i \in \{1, \ldots, 6\}. \tag{2.9}$$

Since precordial leads share the same reference and observe neighbouring thoracic positions, they are not mutually independent.

A permutation involving only RA, LA and LL does not change the WCT, because their sum remain unchanged, and therefore, there would be an interchange among the three limb leads, but the common reference shared by the precordial leads would remain untouched. In contrast, an interchange involving RL can replace one of the potentials included in Equation 2.8, thus modifying the WTC and affecting all six precordial leads.

That is why a correctly acquired ECG does not depend only on the individual correctness of the lead, but also on the consistency between the frontal plane electrical axis, the morphology of the limb leads and the progression of the precordial leads.

2.1.3 Cardiac Electrical Axis and Inter-Subject Variability

The cardiac electrical axis is the average direction of ventricular depolarization in the frontal plane. It can be estimated from the limb leads and it is expressed as an angle, which in adults is approximately between $-30°$ and $+90°$.

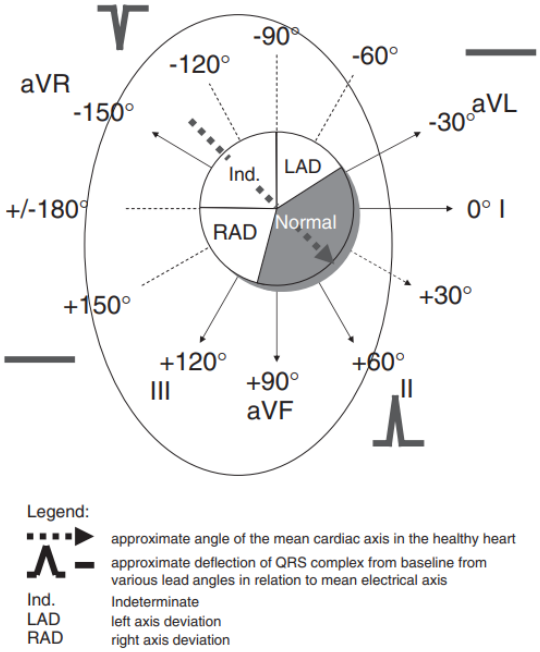


Figure 2.5: Schematic representation of the cardiac electrical axis in the hexaxial reference system Kuhn and Rose [2008].

Its clinical relevance comes from the fact that a deviated electrical axis may be associated with different pathologies, like ventricular hypertrophy, conduction disorders and myocardial infarction. The electrical axis is also relevant in this work because limb electrode interchange may produce an apparent axis deviation, due to alterations in polarity and identity of the frontal leads. Also, it is fair to say that an abnormal electrical axis is not alone sufficient to detect an acquisition error.

Furthermore, the heart electrical axis is affected by anatomical variability. The position and orientation of the heart may vary according to body composition, thoracic anatomy, sex, age and the presence of cardiovascular conditions. Kayyali et al. reported a relationship between anatomical cardiac orientation and the corresponding electrical axis, highlighting the influence of individual characteristics such as body mass index and sex Kayyali et al. [2025].

This represents another challenge in the context of this work, because an automatic detection system should also be able to distinguish between acquisition errors and deviations in the electrical axis due to anatomical reasons.

2.1.4 Standard Electrode Placement

In a standard ECG, the RA and LA electrodes are positioned on the right and left arms, while RL and LL are positioned on the right and left legs. In special occasions, such as exercise testing or ambulatory monitoring, the limb leads could be placed at proximal sites, such as wrists and ankles, or on the torso. Any non standard configuration should be clearly documented.

The six precordial electrodes should be placed in specific anatomical positions Kligfield et al. [2007]:

- V_1: fourth intercostal space at the right sternal border;
- V_2: fourth intercostal space at the left sternal border;
- V_3: midway between V_2 and V_4;
- V_4: fifth intercostal space at the left midclavicular line;
- V_5: left anterior axillary line, horizontally aligned with V_4;
- V_6: left mid-axillary line, horizontally aligned with V_4 and V_5.

This represents another problem in this context because to correctly position the precordial electrodes in the torso, these specific anatomical positions should be correctly identified, and this is not always straightforward. For example, the sternal angle is generally used to locate the second rib and to count the intercostal

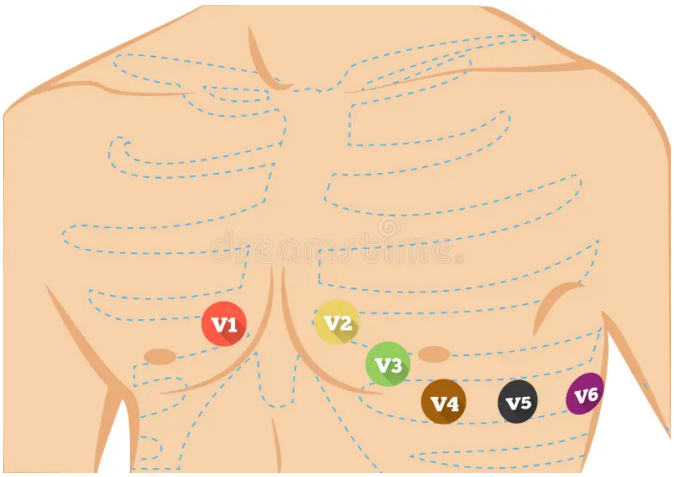


Figure 2.6: Standard anatomical placement of the six precordial electrodes Dreamstime [2026].

spaces. However, operators could find palpation difficult due to different anatomical factors, such as obesity, muscular development and breast tissue. A frequent error committed by operator during the acquisition is to place V_1 and V_2 one or more intercostal spaces above their supposed positions. This may alter the signal morphology, generating patterns resembling right bundle branch block, anterior ischaemia, or Brugada-like morphology Ilg and Lehmann [2012], García-Niebla et al. [2014], Kania et al. [2020].

2.2 ECG Acquisition Errors

The present work focuses on the detection of *electrode interchange*, which is the incorrect association of two or more electrodes or cables with their expected recording channels or anatomical sites. In this scenario, signals that are acquired from valid electrode positions are assigned to incorrect channels. This acquisition error should be distinguished from *anatomical misplacement*, in which an electrode is positioned in its correct anatomical position, but with an offset. Since the misplacement produces a signal acquired from non standard positions, it cannot be seen as a simple permutation of the standard ECG channel. Anatomical misplacement is mentioned for completeness, but it is not addressed in the experimental section of this thesis.

This distinction is clear in the precordial electrodes. For example, an interchange between V_1 and V_2 corresponds to the exchange of the two recorded channels. In contrast, positioning V_1 in the second or third intercostal space produces a signal that originates from a non standard anatomical position and therefore, it could not correspond to any of the six precordial recordings.

2.2.1 Limb Electrode Interchange

Considering four limb electrode labels, the number of possible assignments to the four anatomical positions is:

$$4! = 24. \tag{2.10}$$

Only one is correct, so this leaves 23 mathematically incorrect permutations. These permutations do not necessarily produce 23 distinct ECG patterns because some assignments are easily recognizable. For example, RL-related exchanges often produce near-isoelectric leads or abnormal amplitudes. All of the 13 representative configurations are illustrated in Figures 2.7 and 2.8.

Since RL is not used directly to construct the bipolar leads I, II, and III, its interchange with an active limb electrode can have different consequences. In some cases, interchanges involving the RL electrode produce a characteristic nearly

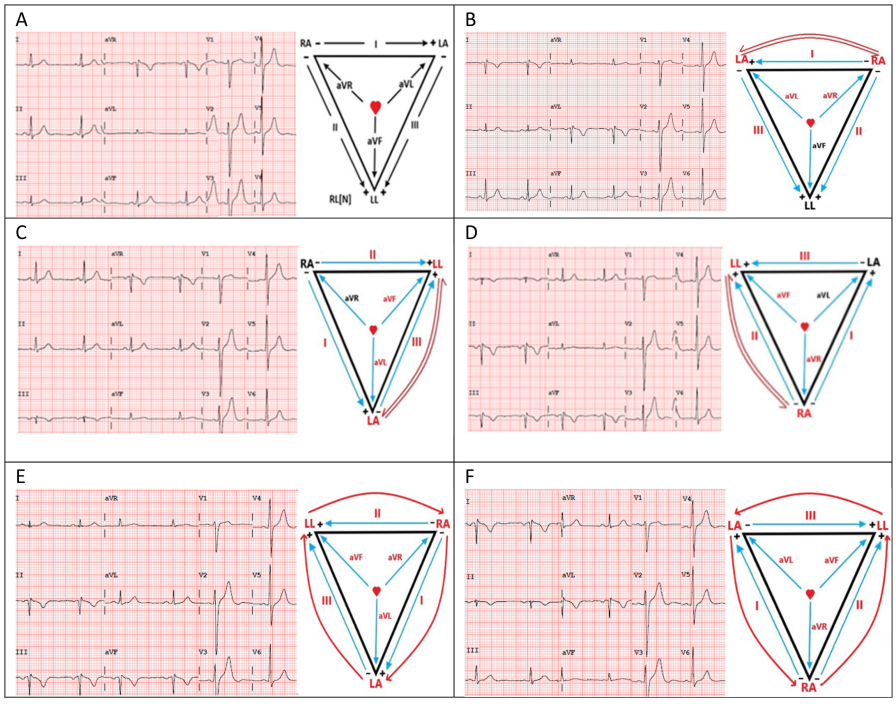


Figure 2.7: Effects of selected limb-electrode interchanges not involving the RL electrode Batchvarov et al. [2007].

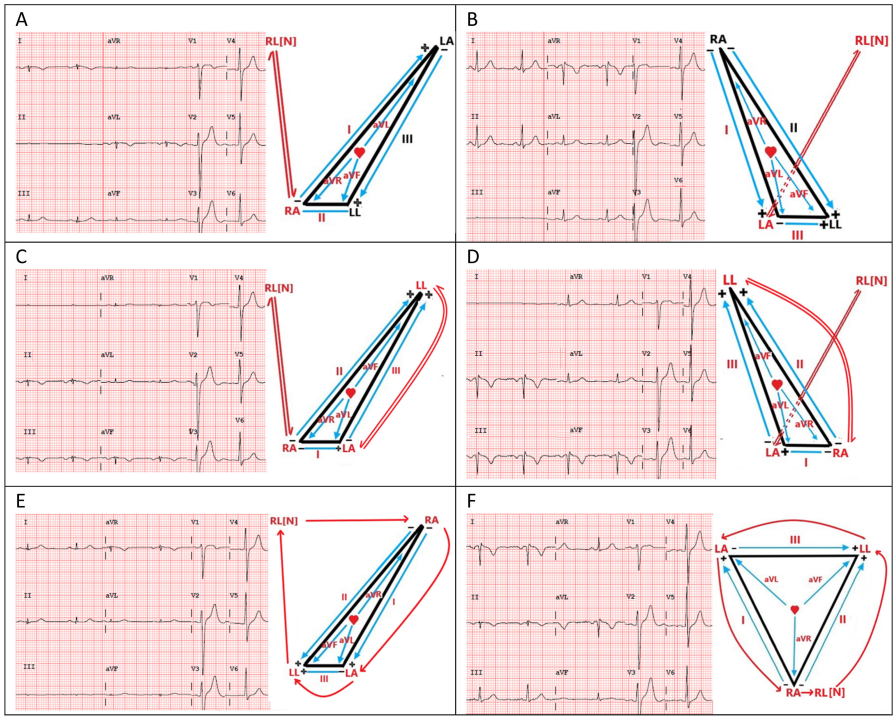


Figure 2.8: Effects of selected limb-electrode interchanges involving the RL electrode Batchvarov et al. [2007].

isoelectric lead and can therefore be easily recognised. However, other configurations may produce limited changes or be indistinguishable from other electrode interchange classes:

- RA–RL interchange produces a nearly flat lead II, as shown in Figure 2.8A;
- LA–RL interchange produces a nearly flat lead III, as shown in Figure 2.8B;
- the combined LA–LL and RA–RL interchange produces a nearly flat lead I, as shown in Figure 2.8C;
- LL–RL interchange does not produce appreciable changes with respect to the normal configuration, as shown in Figure 2.7A;
- the combined LL–RA and RL–LA interchange produces a nearly flat lead I, as shown in Figure 2.8D;
- the clockwise cyclic configuration involving RL produces a nearly flat lead III, as shown in Figure 2.8E;
- the counter-clockwise cyclic configuration involving RL produces the same lead pattern as the corresponding counter-clockwise rotation without RL, because the additional LL–RL interchange does not modify the resulting configuration. The two cases are shown in Figures 2.8F and 2.7F, respectively.

Not all configurations including RL are distinguishable among each other, and therefore they are not suitable for being separate classes. The limb electrode classification problem analysed in this thesis includes the normal configuration and five abnormal patterns: three pairwise interchanges among RA, LA, and LL, together with clockwise and counter-clockwise cyclic rotations of the three active electrodes. The considered classes are present in Figure 2.7

The transformations described below are expressed in terms of the original limb leads I, II, III, aVR, aVL, and aVF. A prime denotes the lead obtained after the corresponding electrode interchange.

LA-RA Interchange

LA-RA interchange occurs when the left arm and right arm electrodes are interchanged. In this case, the transformed limb leads are:

$$I' = -I, \tag{2.11}$$
$$II' = III, \tag{2.12}$$
$$III' = II, \tag{2.13}$$
$$aVR' = aVL, \tag{2.14}$$
$$aVL' = aVR, \tag{2.15}$$
$$aVF' = aVF. \tag{2.16}$$

So, lead I is inverted, leads II and III are exchanged, aVR and aVL are exchanged, and aVF remains unchanged. From a geometric point of view, this configuration can be interpreted as a reflection of Einthoven's triangle about the aVF axis Burns and Buttner [2024], Batchvarov et al. [2007].

From a clinical point of view, this condition may resemble dextrocardia, but unlike what happens when this condition is truly present, the normal precordial R-wave progression is generally preserved because the precordial electrodes are not involved in the interchange Batchvarov et al. [2007].

RA-LL Interchange

RA-LL interchange is the interchange of the right arm and left leg electrodes. The resulting transformations are:

$$I' = -III, \tag{2.17}$$
$$II' = -II, \tag{2.18}$$
$$III' = -I, \tag{2.19}$$
$$aVR' = aVF, \tag{2.20}$$
$$aVL' = aVL, \tag{2.21}$$
$$aVF' = aVR. \tag{2.22}$$

In this configuration, lead II is inverted, leads I and III are exchanged and inverted, aVR and aVF are exchanged, and aVL remains unchanged. The transformation can be interpreted as a rotation of Einthoven's triangle about the aVL axis Burns and Buttner [2024], Batchvarov et al. [2007].

The pattern resulting from this interchange may resemble an inferior myocardial infarction and may also suggest an abnormal atrial activation sequence. What distinguishes this error from a genuine pathological condition is the fact that in this pattern lead aVL is preserved, together with normal precordial progression Batchvarov et al. [2007].

LA-LL Interchange

In the LA-LL interchange, the left arm and left leg electrodes are interchanged. The transformed leads are:

$$I' = II, \tag{2.23}$$
$$II' = I, \tag{2.24}$$
$$III' = -III, \tag{2.25}$$
$$aVR' = aVR, \tag{2.26}$$
$$aVL' = aVF, \tag{2.27}$$
$$aVF' = aVL. \tag{2.28}$$

Leads I and II are exchanged, lead III is inverted, aVL and aVF are exchanged, and aVR remains unchanged. This configuration corresponds to a transformation of Einthoven's triangle on the aVR axis Burns and Buttner [2024], Batchvarov et al. [2007].

This class of interchange may produce an apparent inferior ischaemic or infarction-related pattern.

Clockwise Rotation

The three active electrodes could also be wrongly assigned in a cyclic permutation. In this work, the class `ROT_ORA` corresponds to the clockwise rotation defined by:

$$RA' = LA, \qquad LA' = LL, \qquad LL' = RA. \tag{2.29}$$

The corresponding limb-lead transformations are:

$$I' = III, \tag{2.30}$$
$$II' = -I, \tag{2.31}$$
$$III' = -II, \tag{2.32}$$
$$aVR' = aVL, \tag{2.33}$$
$$aVL' = aVF, \tag{2.34}$$
$$aVF' = aVR. \tag{2.35}$$

The augmented leads follow this sequence:

$$aVR \to aVF, \qquad aVF \to aVL, \qquad aVL \to aVR, \tag{2.36}$$

Counter-Clockwise Rotation

The three active electrode could also be rotated in the opposite direction. This class is denoted as `ROT_ANT`, and it is defined by:

$$RA' = LL, \qquad LA' = RA, \qquad LL' = LA. \tag{2.37}$$

The transformed leads are:

$$I' = -II, \tag{2.38}$$
$$II' = -III, \tag{2.39}$$
$$III' = I, \tag{2.40}$$
$$aVR' = aVF, \tag{2.41}$$
$$aVL' = aVR, \tag{2.42}$$
$$aVF' = aVL. \tag{2.43}$$

2.2.2 Precordial Electrode Interchange

A precordial electrode interchange occurs when the signals acquired by two thoracic electrodes are associated with the wrong recording channels.

For six precordial electrodes, the number of unique pairwise interchanges is:

$$\binom{6}{2} = \frac{6!}{2!\,4!} = 15. \tag{2.44}$$

So, a complete pairwise classification problem contains 16 classes, where the added one is the correct configuration. Simulating a precordial interchange is straightforward, but detecting it can be challenging. This is because adjacent precordial electrodes observe thorax regions that are close to each other and, therefore that may have similar amplitudes and morphologies.

Interchanges involving more distant leads may produce more evident errors. However, detection of acquisition error is always challenged by the presence of pathological conditions, such as myocardial infarction, ventricular hypertrophy, conduction disorders, or poor R-wave progression, which may alter the expected precordial sequence.

2.3 Automatic Detection of Electrode Errors

The objective of this section is to describe the evolution over time of the methodologies used to automatically detect electrode errors.

Earliest approaches relied more on the morphology of the signal, for example, the expected ECG polarities, amplitudes and lead relationships. With the spreading of machine learning approaches, later methods introduced reconstructed leads, statistical descriptors and supervised classifiers. The most recent transition towards automatic feature learning is deep learning approaches. In general, in literature, limb and precordial interchanges are addressed as two separate and different tasks, since limb leads have known algebraic relationships, while precordial leads are mainly characterized by spatial progression and inter-lead similarity. Clarification of the terminology is needed: some research papers use *lead reversal*, *lead-wire interchange*, or *lead misplacement*. The present work uses *electrode interchange*.

The main methodological families are:

- Rule-Based and Reconstruction Methods
- Feature Engineering and Traditional Machine Learning
- Deep Learning Approaches

2.3.1 Rule-Based and Inter-Lead Correlation Methods

The earliest approaches relied on rule-based methods, which compare the observed ECG with known physiological patterns. Typical indicators include unexpected P-wave orientation, inversion of normally positive leads, unexpected positivity of aVR, unusually low or nearly flat signal amplitude, abnormal QRS amplitude relationships. These rules are easy to interpret and computationally inexpensive Rjoob et al. [2020], Huang et al. [2024].

A single one of these rules is not sufficient to detect an interchange, because negative or unusual waveforms may also result from abnormal electrical axis, conduction disorders, ectopic rhythms, pathological morphology and poor electrode contact. Since these rules are fixed, they may not be able to account for inter-patient variability. Furthermore, a rule developed for one interchange may not identify other configurations. That is why this approach is best suited for well defined configurations that produce clearly recognizable ECG alterations Rjoob et al. [2020], Huang et al. [2024].

Another approach is the inter lead correlation, which exploits relationships between channels. This is useful because the leads exhibit similarities, which may be altered by electrode interchange. Jekova et al. investigated the inter-lead correlation analysis for the automated detection of cable reversals in standard and extended ECG recordings Jekova et al. [2016]. The authors calculated a correlation coefficient that describes the similarity between leads to determine whether the observed configuration was compatible with the expected channel ordering.

Correlation methods can analyse relationships between several channels with limited computational cost, but it is important to note that correlation alone may be insufficient for difficult error configuration, since correlation between leads can be influenced by different factors, such as noise and pathological morphology Jekova et al. [2016], Huang et al. [2024].

Although these methods are computationally accessible and interpretable, their performance depends on predefined criteria. This motivated the development of techniques based on feature engineering and machine learning.

2.3.2 Feature Engineering and Traditional Machine Learning

Another method for the automatic detection of electrode errors is to use traditional machine learning approaches. Machine learning models learn a decision function from a set of labelled ECG examples. However, these models do not operate directly on the ECG waveform, but first the ECG must be converted into numerical feature vectors which contain the selected characteristics of the signal.

Once the feature vector is ready, it is given as input to a classifier. As reviewed by Rjoob et al., the most common approaches proposed in the literature include support vector machines, random forests and other supervised classifiers Rjoob et al. [2020]. These techniques can combine several signal characteristics within a single classification model and are therefore more flexible than single decision rules.

This approach requires a manual feature engineering procedure, in which characteristics considered relevant to the classification task are identified and extracted before model training. Therefore, the effectiveness of the classifier depends not only on the selected learning algorithms, but also on the ability of the chosen features to represent the alterations produced by electrode interchange.

For this reason, in recent years there was a transition towards deep learning models that learn relevant representations directly from ECG signals Huang et al. [2024].

2.3.3 Deep Learning Approaches

Deep learning approaches provide an alternative to rule-based methods and traditional machine learning pipelines. Neural networks can directly process ECG signals and learn the representation that is most relevant to the classification task. In the context of this work, this approach makes it possible to identify patterns associated with electrode errors without a previous step of feature extraction Huang et al. [2024].

A relevant development is represented by *lightweight* deep learning architectures, which are designed to limit the number of trainable parameters and, therefore, the associated computational cost, while preserving the ability to analyse ECG signals. This is particularly relevant in the context of telemedicine and portable acquisition devices Huang et al. [2024]. Portable systems generally have more limited memory and resources than conventional workstations, so the trade-off between model complexity and predictive performance is an important aspect in the design of automatic ECG validation systems.

Among the lightweight approaches proposed for electrode error detection, the models introduced by Huang et al. are particularly relevant to the present work Huang et al. [2024]. The authors proposed two architectures for the two separate classification tasks: LDenseNet was developed for the limb electrode error detection task, while the Inter Lead Correlation model was designed for the precordial case.

In the original study, Huang et al. designed the LDenseNet to distinguish between the normal configuration and three limb electrode interchanges: LA-RA, RA-LL and LA-LL. For the precordial task, the Inter Lead Correlation model considered the normal configuration and nine pairwise interchanges. Limb electrode errors were generated by mathematical lead transformations, while the precordial interchanges were simulated by exchanging the corresponding channels Huang et al. [2024].

These models demonstrated the ability of performing multi-class electrode error detection using architectures with a limited number of trainable parameters. The study conducted by Huang et al. was based on simulated interchange configurations and the authors reported that poor signal quality of ECG recordings contributed to false positive predictions Huang et al. [2024]. This underlines the importance of combining electrode interchange classification with appropriate preprocessing and signal quality analysis.

These two architectures proposed by Huang et al. represent the main reference for this thesis. The implemented architectures with the adaptations taken from Huang et al. are described in detail in Chapter 3.

2.4 Data Simulation and Signal Quality in Previous Studies

2.4.1 Simulation of Electrode Interchanges

Electrode interchange detection using supervised models can be challenging due to the limited availability of clinically verified ECG records that contain known

interchange configurations. In real ECG databases, the records that contain acquisition errors could not be present because they were discarded or could be present in a number that is insufficient to train a multi-class classifier. For this reason, previous studies have generated interchange configurations from correctly acquired ECG recordings Huang et al. [2024].

For limb electrode interchange, the simulated examples can be obtained through deterministic transformations derived from the mathematical definitions of the limb leads. This approach makes it possible to generate labelled examples of specific limb electrode configurations in a controlled manner Huang et al. [2024].

Precordial interchange can be simulated directly by exchanging two channels among $V_1, \ldots, V_6$, while leaving the other channels untouched. This is the approach used by Huang et al. to simulate the precordial classes considered in their study Huang et al. [2024].

Simulation offers several advantages: it provides records with exact class labels, it allows to examine different interchange configurations starting from the same patient, and also, it facilitates the construction of balanced training sets. On the other hand, simulated records represent some challenges. Real acquisitions errors occur together with motion artifacts, poor electrode contact, baseline wander or other forms of signal degradation that cannot be reproduced by a simple channel transformation.

It is important to underline that simulated electrode interchange should not be confused with anatomical electrode misplacement. In an interchange the signals are acquired from standard electrode positions, but they are wrongly reassigned, while in an anatomical misplacement, the signal is produced from a non standard body location. The experimental work presented in this thesis focuses on electrode interchange and does not address anatomical misplacement.

2.4.2 Signal Quality Assessment and Preprocessing

A very important stage in the context of this work is signal preprocessing. In automatic ECG analysis, acquisition artifacts can alter the morphology of the signal, therefore, the objective of preprocessing is to attenuate components that are unrelated to cardiac activity, while preserving the characteristics that are clinically relevant, such as the P wave, QRS complex, and T wave Satija et al. [2018].

Common preprocessing operations include the attenuation of the baseline wander, high frequency noise and electrical interference, along with resampling and normalization.

It is important to note that preprocessing alone is not a sufficient step for automatic electrode interchange analysis. Signal Quality Assessment evaluates the usability of an ECG through numerical criteria commonly referred to as Signal Quality Indexes. In the approach described by Xie et al., quality assessment

is based on statistical and morphological information, including measures such as kurtosis, skewness, and consistency among leads Xie et al. [2021]. These indicators are useful to determine if an ECG should be kept or rejected before further analysis.

Signal quality assessment should be considered a distinct task with respect to the electrode interchange detection task. Indeed, a technically clean ECG may still contain an electrode interchange, while a correctly acquired ECG may be affected by noise or poor electrode contact. Furthermore, poor quality of an ECG record may modify the same morphological and inter lead relationships used to recognize interchange, and may therefore lead to wrong model predictions.

The correlation between poor signal quality and false positive predictions was also observed by Huang et al. in their studies. Therefore, the authors identified the combination of electrode interchange detection with an ECG quality assessment algorithm as a possible winning strategy Huang et al. [2024].

Consequently, the signal quality validation step should be seen as a preliminary stage, whose purpose is not to formulate a diagnosis, but to verify that the signal contains sufficient and readable technical information for the subsequent classification task.

2.5 Research Gaps and Positioning of the Thesis

Even though the electrode interchange detection analysis progressed extensively, several limitations remain. For example, the previously cited problem of the scarcity of clinically verified ECG records containing electrode interchange configurations. For this reason, supervised approaches use simulated configurations, which however, do not fully reproduce the variability encountered during real ECG acquisitions Rjoob et al. [2020], Huang et al. [2024].

Another limitation of the previous studies is the number of error classes covered. For example, the original models proposed by Huang et al., from which this work was inspired, considered three pairwise interchanges for the limb electrodes and nine interchange configurations for the precordial ones. Therefore, neither the cyclic permutations of the limb electrodes nor all fifteen possible pairwise precordial interchanges were included in the original classification tasks Huang et al. [2024].

Previous approaches also tend to address signal quality and electrode interchange detection as separate problems. However, these two aspects interact with each other because poor quality ECG may generate unreliable interchange predictions. Therefore, a complete pipeline should include a signal quality assessment step prior to the interchange classification task.

The present thesis is collocated in this context and it tries to address all these limitations mentioned before. It adapts to the architectures proposed by Huang et al. Huang et al. [2024], by including more interchange classes; the limb-electrode

task includes the normal configuration, the three pairwise interchanges LA -RA, RA–LL, and LA–LL, and the two possible cyclic rotations of the active limb electrodes. The precordial task includes the normal configuration and all fifteen pairwise interchanges among $V_1, \ldots, V_6$.

Furthermore, this work includes a preliminary step that includes signal preprocessing and signal quality assessment. All about datasets, simulation procedures, preprocessing steps, adapted architectures, training strategy, and evaluation protocol are widely described in Chapter 3.

Chapter 3

Methods and Proposed Solution

3.1 Signal Preprocessing Pipeline

Digital ECG signals cannot be directly used as input to neural networks, because the acquisition chain introduces several artifacts, such as 50 Hz power-line interference and baseline wander due to respiration or patient movement Redmond et al. [2008]. Moreover, the native sampling rate of the devices is often higher than what is required to preserve the clinically relevant content of interest. In addition, there is a variability in signal amplitude between different recordings, caused by differences in electrode impedance, device gain and patient physiology. As a consequence, a normalization phase is required to make the samples statistically comparable.

In the specific context of this work, an additional challenge is related to the use of Holter monitors. Compared to standard clinical ECGs, these devices acquire recordings of much longer duration and are more exposed to muscle noise and motion artifacts Satija et al. [2018]. Therefore, the signal quality thresholds had to be calibrated with particular care.

This chapter describes the different stages of the preprocessing pipeline adopted in this thesis, in order to obtain a clean filtered signal of enough quality to be processed by the network.

3.1.1 Raw Signal Structure, Quality Control and Label Normalization

The analysis of each ECG recording starts from the *raw* signal, that is, the time sequence of samples as acquired by the device, before segmentation and transformation. The original signal is sampled at a frequency of 1 kHz; later, within the preprocessing pipeline, it can be resampled at a lower frequency, in order to reduce the computational cost of downstream stages and to standardize the handling of time windows.

Structurally, a recording is represented as an associative structure in which each lead is associated with its corresponding sampled sequence. In abstract form, the record can be described as

$$\mathbf{X} = \{x_\ell[n]\}_{\ell \in \mathcal{L}}, \tag{3.1}$$

where $\mathcal{L}$ denotes the set of available leads and $x_\ell[n]$ is the value of lead ℓ at discrete sample n. In this work, only recordings that contain all 12 standard leads are passed to the quality control stage; after the preliminary checks, every ECG is therefore treated as a synchronized 12-channel multilead signal.

The 12 leads considered correspond to the standard international configuration, composed of six limb leads (`I`, `II`, `III`, `aVr`, `aVl`, `aVf`) and six precordial leads (`V1`–`V6`). To standardize the internal representation of the signal, these leads are ordered in the following canonical sequence:

$$\texttt{ALL_LEADS} = [I, II, III, aVr, aVl, aVf, V1, V2, V3, V4, V5, V6]. \tag{3.2}$$

This ordering is enforced at implementation level and is preserved in all subsequent stages of the pipeline, ensuring consistency between signal reading, visualization, quality control and input construction for the model.

During EDF file reading, the original lead labels are first cleaned by removing prefixes and non-informative components such as `ECG`, `LEAD` and the "-" character. Then, any nomenclature variants are mapped to a unified convention consistent with the notation used in this thesis. Specifically, the forms `AVR`, `AVL`, `AVF`, `DI`, `DII` and `DIII` are converted to `aVr`, `aVl`, `aVf`, `I`, `II` and `III`, respectively. In this way, even in the presence of heterogeneous labeling in the source files, the signal is always returned with a coherent nomenclature and with the predefined lead ordering.

From an implementation point of view, the record is represented as a dictionary containing the set of signals and a metadata field describing the total duration of the trace. In simplified form, the structure is:

```
{
  'signals': {
    'I': array([...]),
```

```
    'II': array([...]),
    'III': array([...]),
    'aVr': array([...]),
    'aVl': array([...]),
    'aVf': array([...]),
    'V1': array([...]),
    ...
  },
  'duration_sec': ...
}
```

In this representation, each key of the dictionary `signals` identifies one lead, while the associated value contains the corresponding vector of samples.

The duration of the trace is not assumed to be constant *a priori*. Although the pipeline uses a nominal length of 30 s as an operational reference, not all available recordings match this exact time span. For this reason, in the preliminary stages, filters are applied to exclude traces that are too short, incomplete or otherwise unsuitable for analysis, in order to preserve a greater homogeneity among the signals that are actually processed.

The pipeline also includes visualization tools for qualitative inspection of the signal. In particular, there are functions to display the full recording over all 12 leads in the canonical order, as well as functions to visualize limited portions of the trace on a selected subset of leads. In addition, a dedicated procedure automatically generates plots for discarded signals, visually highlighting problematic leads and facilitating the analysis of exclusion causes. Figure 3.1 shows an example of the raw ECG printout, which is useful for connecting the numerical representation of the signal to its corresponding clinical visualization. The figure illustrates how the 12 leads are presented in the conventional ECG report format and provides an important interpretative support for understanding the relationship between data structure, channel layout and signal morphology.

3.1.2 Power-Line Removal, Digital Filtering and Polyphase Resampling

The preprocessing pipeline is defined as a sequence of operations applied to each ECG lead. The goal is to reduce noise and to standardize signal characteristics before they are used as input to the neural network.

As a preliminary step, the signal is converted to a homogeneous numeric format, and any invalid values (NaN or infinite) are replaced with zeros. This operation ensures numerical stability in the subsequent filtering stages.

The first preprocessing operation is the removal of power-line interference using a notch filter centered at 50 Hz. Power-line interference is one of the main sources

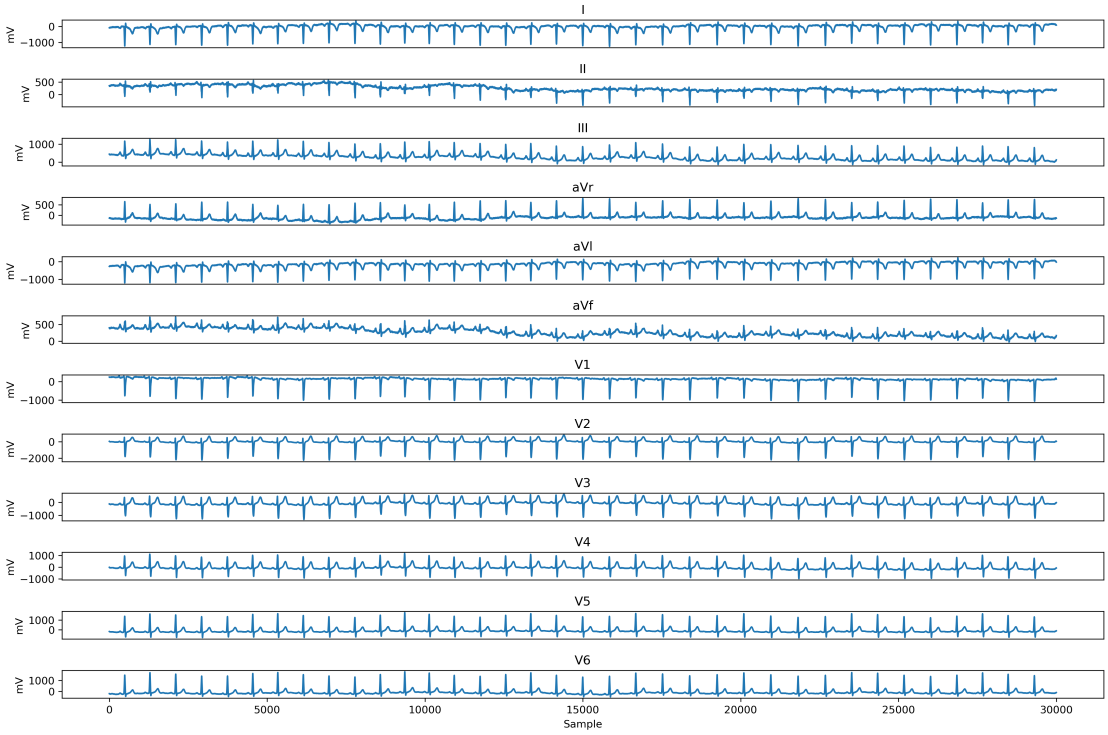


Figure 3.1: Example of *raw* 30-second ECG printout with the 12 leads ordered according to the convention adopted in the pipeline.

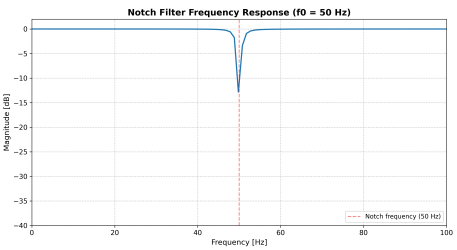
of noise in ECG signals acquired in clinical environments and appears as a quasi-constant sinusoidal component at 50 Hz Redmond et al. [2008]. If not removed, it can degrade signal quality and affect the following stages of the pipeline (see Figure 3.2). The notch filter is implemented using the function `iirnotch()` from the SciPy library, with a central frequency of 50 Hz and a quality factor $Q = 30$. This parameter controls the selectivity of the filter, that is, the width of the attenuated band around the central frequency. Once the filter coefficients are computed, the filter is applied using the function `filtfilt()` from the SciPy library, which performs forward and backward filtering in time. In this way, the overall filtering has zero phase Redmond et al. [2008], and the temporal position of the main ECG components, such as the QRS complex and the P and T waves, is preserved, avoiding distortions that could negatively impact later processing stages.

The second stage of the pipeline consists of a band-pass filter with cutoff frequencies of 0.5 Hz and 120 Hz Satija et al. [2018] (see Figure 3.3). This step aims to preserve the frequency band that contains clinically relevant ECG information, while attenuating less relevant components at very low and very high frequencies. The lower cutoff frequency (0.5 Hz) attenuates slow baseline fluctuations (baseline wander), typically caused by respiration or patient movement, while preserving low-frequency diagnostic components such as the T wave and the ST segment Satija et al. [2018]. The upper cutoff frequency is set to 120 Hz. Standard clinical ECGs are often filtered with upper limits between 80 and 100 Hz, but in this work a higher value is used to avoid excessive attenuation of the high-frequency components of the QRS complex. This choice is particularly useful for Holter recordings, where beat morphology shows greater variability. The filter is implemented in second-order sections (Second-Order Sections, SOS) form, which improves numerical stability compared to direct-form realizations and reduces rounding errors typical of high-order IIR filters. The band-pass filtering is also applied in a zero-phase manner.

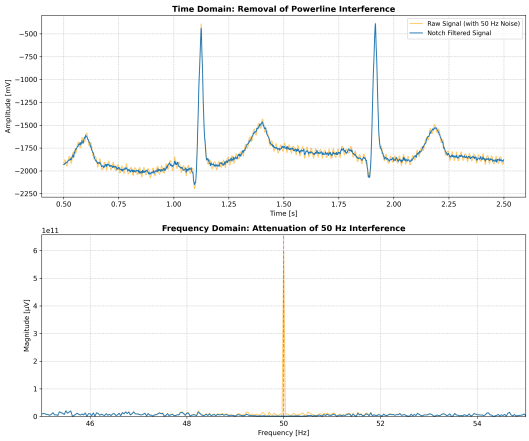
The last stage of the pipeline is the resampling of the signal from the original sampling frequency of 1000 Hz to 250 Hz (see Figure 3.4). Since the final sampling rate is lower than the initial one, this operation effectively reduces the number of samples by a factor of four. In contrast to simple decimation, the polyphase resampling approach internally applies an anti-aliasing FIR filter that limits aliasing of frequency components above the new Nyquist frequency of 125 Hz. The band-pass filter applied in the previous step already attenuates spectral content beyond 120 Hz, so that almost all useful information is confined below the new Nyquist frequency. As a result, resampling can be performed without introducing significant aliasing phenomena. The reduction in sampling rate decreases the number of samples and the computational cost of subsequent processing stages, while preserving the clinically relevant information contained in the ECG signal.

At the end of the pipeline, an additional consistency check replaces any remaining non-finite values with zeros. Although such values are not expected under normal operation of the filtering and resampling algorithms, this operation is a robustness measure against possible anomalous signal conditions. Finally, the signal is converted to the `float32` format, reducing memory consumption and making the data directly usable in the deep learning libraries adopted for training.

To visualize the effectiveness of the filtering and resampling operations, Figures 3.2, 3.3 and 3.4 report representative plots of the action of each algorithmic block.

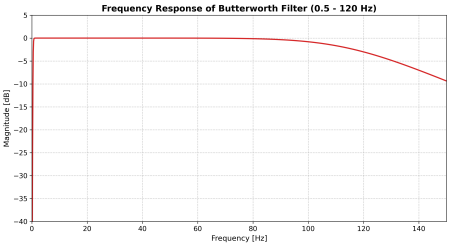


Frequency response of the notch filter.

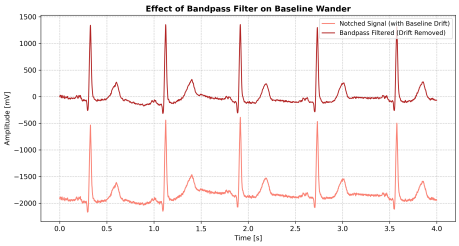


ECG before and after notch filtering.

Figure 3.2: Effect of the 50 Hz notch filter on power-line interference.



Frequency response of the band-pass filter.



ECG before and after band-pass filtering.

Figure 3.3: Effect of the band-pass filter (0.5–120 Hz) on baseline drift.

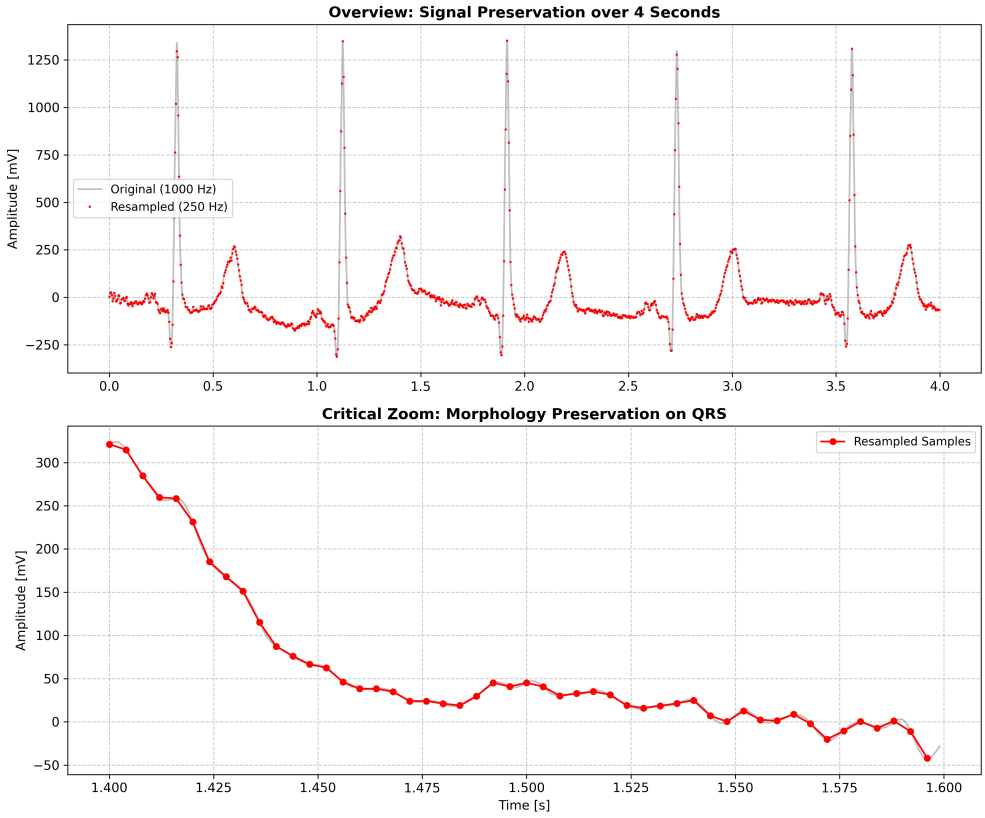


Figure 3.4: Effect of polyphase resampling: comparison between the original signal (1000 Hz) and the resampled signal (250 Hz).

3.1.3 Signal Quality Assessment (SQA) and Structural Detection of the Neutral Electrode

While the filtering stage described above aims to clean the signal from unwanted components such as power-line interference and baseline drift, the signal quality assessment (SQA) stage selects those signals that, after filtering, still contain useful cardiac information and can therefore be used for training the neural network Satija et al. [2018]. This step is necessary because an ECG can be technically recorded without saturation and with proper sampling, but still contain unusable leads, for example flat lines due to a detached electrode or signals with almost zero amplitude caused by an incorrect connection of the reference cable Rjoob et al. [2020].

To maximize the amount of data available for training and to provide homogeneous segments to the model, the proposed architecture splits continuous ECG recordings into many shorter time windows and applies amplitude normalization to make the samples statistically comparable. It is important to note that SQA is deliberately performed before segmentation and normalization, immediately after filtering. This design choice is motivated by the need to run quality checks on the signal expressed in its original physical units (μV), so that thresholds can be interpreted as physically and clinically meaningful quantities. If the SQA block were applied after normalization, the rejection thresholds would need to be recalibrated on an arbitrary and dimensionless scale.

Since the dataset includes both standard 30-second ECGs and longer Holter recordings, which are noisier and more affected by baseline drift, the pipeline implements two dynamic configurations. Based on the metadata of each recording, the algorithm selects stricter thresholds for standard ECGs and slightly more tolerant thresholds for Holter signals, in order not to discard clinically valid traces Satija et al. [2018]. Moreover, thresholds differ between limb and precordial leads, to reflect their different amplitude physiology Huang et al. [2024].

The signal quality assessment is organized into three levels:

- **Level 1: Global lead check.** For each lead, global statistics are computed: standard deviation (σ), peak-to-peak amplitude (ptp), median of the absolute value (median($|x|$)), and median absolute deviation of the first derivative (MAD(Δx)). These values are compared with fixed thresholds to detect macroscopic conditions such as flat signals, ADC saturation, or complete absence of cardiac morphology Huang et al. [2024]. If the lead fails this level, it is discarded immediately.

- **Level 2: Window-wise check.** If the global check is passed, the signal is split into 2-second windows (500 samples at 250 Hz). A window length of 2 seconds was chosen as a compromise between temporal resolution and morphological completeness: for physiological heart rates in the 30–300$\tilde{\text{b}}$pm

range, it typically contains at least one QRS complex together with its surrounding P and T waves, while remaining short enough to track rapid changes in signal quality over time. For each window, the same statistics as above are computed, together with a check that verifies the presence of at least one QRS complex with a physiologically plausible heart rate (30–300 bpm) Huang et al. [2024]. A window is marked as invalid if at least one of the following criteria is satisfied; otherwise, it is considered valid:

Table 3.1: Invalidation criteria for time windows at Level 2 of the Signal Quality Assessment.

Check	**Condition**	**Disturbance detected**
Flatline	$\sigma < 25\,\mu$V and $ptp < 40\,\mu$V	Detached electrode
Low energy	$\sigma < 15\,\mu$V (limb leads)	Signal too weak
Low amplitude	median$(\lvert x\rvert) < 14\,\mu$V	Absence of potential (e.g., wrong RL)
No morphology	MAD$(\Delta x) < 1.5\,\mu$V	Absence of QRS cardiac activity
Noise	MAD$(\Delta x) > 25\,\mu$V	EMG noise or motion artefact
Clipping	Sat. $> 2\%$ samples	ADC saturation
ADC step	max$(\lvert\Delta x\rvert) > 2000\,\mu$V	Anomalous ADC jump
Baseline drift	$\lvert$mean$(x)\rvert > 600\,\mu$V	Strong baseline fluctuation
No heartbeat	No valid QRS peaks	Non-cardiac or unreadable signal

The whole lead is considered valid only if it exceeds a minimum percentage of accepted windows (e.g., 60% for Holter or 70% for standard ECGs). From an implementation perspective, all decision thresholds used in the table are not hard-coded into the algorithm, but centralized in a configuration dictionary. This architectural choice decouples control logic from numeric parameters and makes the SQA pipeline easily scalable and recalibrable for future datasets or specific clinical needs.

- **Level 3: Aggregation at ECG level.** The results are aggregated on the subset of leads considered (either the six limb leads or the six precordial leads, depending on the classification task). A recording is accepted only if at least five out of six leads in the subset are valid. No channel-level removal, replacement, or imputation was performed: when the five-out-of-six criterion was satisfied, the complete multilead window, including all six task-specific leads, was retained. If this condition is not met, the entire ECG is discarded.

Recordings that pass the global checks are then used to build a window mask, that is, a boolean mask indicating, for each non-overlapping 2-second window (stride 2 s), whether the window is locally valid. A window is valid if at least five out of six leads in that time interval are valid. The result is a one-dimensional

vector, which is then applied to the full 12-lead ECG: if a window is discarded, it is discarded simultaneously for all leads. The same window mask computed on the original signal is reused when that signal is involved in the electrode inversion simulations. Using a single validity mask, calculated on the original signal and then applied identically to all synthetic simulations, avoids confusion during training. If window validity were recomputed independently after lead inversion, a misalignment could occur: a window rejected in the normal class might be accepted in a simulated class. With this approach, the only mathematical difference between the normal class and the simulated classes is the spatial alteration of the electrodes.

Most important is that among the SQA checks, a specific role is played by the checks (`low_amplitude` and `no_morphology`), which are useful for detecting recordings containing channel with nearly zero amplitude or without recognizable cardiac morphology. This can happen when an electrode is incorrectly connected to the right leg reference cable, since the reference cable (RL) acts as the ground, so it does not record any cardiac potential, but produces a signal with very low amplitude, though not perfectly flat, because it contains background noise that exceeds the flatline thresholds Jekova et al. [2016]. The `low_amplitude` check ($\text{median}(|x|) < 14\,\mu\text{V}$) detects the systematic near-zero absolute amplitude, while `no_morphology` ($\text{MAD}(\Delta x) < 1.5\,\mu\text{V}$) captures the absence recognizable cardiac morphology Jekova et al. [2016].

It is important to specify that the RL–RA and RL–LA configurations were not included among the target classes of the limb-lead classifier. The final model was defined as a six-class classifier, and samples with labels greater than or equal to 6 were explicitly excluded during data loading for training and evaluation. The SQA criteria were instead used to identify technically unusable recordings and time windows before normalization and model processing.

3.1.4 Normalization via Robust Global Z-Score

Once the signal quality stage is passed, the signals are clinically valid but still show large amplitude variations from patient to patient, due to skin impedance, chest anatomy and acquisition device gain. For this reason, a standardization step is required, so that the neural network receives statistically comparable inputs. However, in this specific scenario, a classical Z-score normalization based on mean and variance has two main issues: sensitivity to outliers and distortion of inter-lead amplitude relationships. To overcome these limitations, a normalization technique called Robust Global Z-Score is implemented.

ECG signals, especially Holter recordings, intrinsically contain high-amplitude artifacts that are rare but numerically extreme. The mean and standard deviation are very sensitive to outliers, so a single anomalous peak can significantly affect

them. For this reason, the median is used instead of the mean to center the signal, and the interquartile range (IQR) is used instead of the standard deviation for scaling Kim et al. [2025]. The IQR is computed as the difference between the 75th and 25th percentiles of the samples Kim et al. [2025]. Under the theoretical assumption of a normal distribution, the IQR can be converted into a robust estimate of the standard deviation by dividing it by a constant factor of approximately 1.34896.

A second problem arises from the specific task of this work, namely the detection of lead inversion. When two electrodes are inverted, the amplitude ratios between leads are distorted: some leads may appear almost flat, while others are amplified Jekova et al. [2016], Rjoob et al. [2020]. If a separate Z-score normalization were applied to each lead, using its local median and local scaling factor, the nearly flat leads would be artificially amplified to achieve unit variance, while the large-amplitude leads would be compressed. The result would be a global flattening that hides the evidence of inversion. For this reason, the IQR is computed globally over the entire subset of leads of interest (either the six limb leads or the six precordial leads).

From a mathematical point of view, the transformation applied to a generic lead $x^{(i)}$ proceeds in two steps:

1. **Computation of the global scaling factor.** Instead of analysing traces separately, all samples from the group of leads under consideration (for example, all six limb or all six precordial leads) are merged to compute a single scaling factor valid for the entire group:

$$\sigma_{\text{global}} = \frac{IQR_{\text{global}}}{1.34896} \tag{3.3}$$

2. **Application of the hybrid normalization.** Each lead is then normalized using local centering and global scaling:

$$x^{(i)}_{\text{norm}} = \frac{x^{(i)} - \text{median}(x^{(i)})}{\sigma_{\text{global}} + \epsilon} \tag{3.4}$$

   In the numerator (local centering), the median of the specific channel is subtracted. This operation recenters the signal around zero and removes slow baseline wander specific to that lead. In the denominator (global division), the signal is divided by the global factor σ_{global}, which is the same for all leads in the group. Dividing by the same value is the key step that preserves the original amplitude ratios between leads. Finally, a small constant $\epsilon = 10^{-8}$ is added to the denominator as a safety measure to avoid division-by-zero errors in the unlikely case that a perfectly flat signal passes the previous quality checks.

3.2 Dataset and Sampling Strategy

This chapter analyses how data are sampled and collected for both training and testing the network. Creating a high-quality dataset is a complex process, as it must be representative of a distribution that realistically mimics the problem under investigation in all its facets; in addition to being balanced and representative, such a distribution must also be sufficiently large to favour generalization.

In Machine Learning, data are extremely important, as they represent the foundations of the entire project, and the ability to correctly manage large amounts of data constitutes a predominant part of a data scientist's work.

In the following, the ECG sampling strategy and the specific creation of different types of datasets are described.

3.2.1 Initial ECG Dataset and Report Organization

The starting point is the large dataset kindly provided by Eng. Manzini, who has collected over time the various electrocardiograms recorded by his device in digital format (EDF). The cardinality of this large set is approximately 180,000 ECG exams, mostly valid, each accompanied by a medical or rejection report.

The index file is a small SQLite database (**records_complete.db**) with the following structure:

- **id** is the exam identifier, from which the file containing the traces can be retrieved;
- **status** can be `"reported"` if the exam has been reported and therefore considered suitable for clinical evaluation, or `"rejected"` if it has been discarded;
- **report** contains the report, encoded in JSON format. The main point is that, by searching within specific classification codes, it is possible to identify both some rejection cases (for example ECG rejected due to unsuitability for reporting, or rejected for administrative reasons and therefore not meaningful for training) and the presence of conditions that may in some way impact the assessment of reportability (for example axis deviation, which may mislead the model). In addition, the JSON contains the indication of the filters that the cardiologist decided to apply, information that is useful for quality assessment.
- **text** is the report text as it is printed. In practice, from the JSON field of the previous column, the set of sentences is extracted, both coded and entered as free text, in order to generate a readable report. This includes comments from the physician, such as suspected electrode inversion.

An example row is the following:

```
{
    "id": 668833,
    "status": "reported",
    "report": {
        "measurements": {
            "HR": [
                {
                    "average": 64
                },
                "bpm"
            ],
            "PQ": [
                {
                    "average": 0.14
                },
                "s"
            ],
            "QRS": [
                {
                    "average": 0.08
                },
                "s"
            ],
            "QT": [
                {
                    "average": 0.36
                },
                "s"
            ],
            "QTc": [
                {
                    "average": 0.372
                },
                "s"
            ]
        },
        "qtcFormula": "BAZETT",
        "adjustments": {
            "timeOffset": 0,
            "filters": [
                {
```

```
                    "id": "4"
                },
                {
                    "id": "5"
                }
            ]
        },
        "codified": [
            {
                "type": "code",
                "value": "BTWSCR01"
            },
            {
                "type": "code",
                "value": "BTWSCA01"
            },
            {
                "type": "code",
                "value": "BTWSCE04"
            },
            {
                "type": "code",
                "value": "BTWSCQ01"
            }
        ]
    },
    "text": "Ritmo sinusale.
             Asse intermedio.
             Non segni elettrici di preciso significato patologico.
             ECG nei limiti della norma."
}
```


The file **report_codes.csv** contains the mapping between the codes provided by the physician and the corresponding text, according to the convention `code/description/category`; representative examples for each category are:

```
BTWSCR01 -> ritmo sinusale -> rhythm
BTWSCX01 -> extrasistoli sopraventricolari -> extrasystoles
BTWSCA01 -> asse intermedio -> axis
BTWSCK01 -> blocco AV di 1° grado -> AVblock
BTWSCB01 -> blocco di branca destro -> blocks
BTWSCV02 -> ipertrofia ventricolare sinistra -> hypertrophy
```

```
BTWSCI03 -> ischemia anteriore -> ischemia
BTWSCF01 -> infarto miocardico anteriore -> infarction
BTWSCE04 -> non segni elettrici di preciso significato patologico ->
↪  recovery_phase
BTWSCQ01 -> ECG nei limiti della norma -> summary
BTWG01 -> rifiutato su richiesta del paziente o richiedente -> rejection
BTWSCVQ2 -> probabile ipertrofia ventricolare -> qualifier
```

3.2.2 Collection, Selection, and Integrity Filters of the Initial Clinical Data

The overall goal is to select a sufficiently large number of electrocardiographic recordings that are not affected by electrode inversion, since these will be simulated later in an artificial way. It is therefore important to carefully select the ECG traces among all those available, so as to be confident that they carry valid information.

To this end, the selection process of "clean" ECGs queries the local database `records_complete.db` and applies the following common filters, which will be reused in the construction of the different datasets:

- **Base filter**: only records with completed status (`status="reported"`) are selected.
- **Manual exclusion (duplicates)**: a list of IDs known a priori to be duplicates (e.g. `'1157262'`, `'1157353'`, etc.) is excluded; this is because, as will be shown later, some normal ECGs may appear in specific test sets and must not be counted multiple times.
- **Textual exclusion (search for artifacts/inversions in the report)**: the free-text part of the report is analysed (converted to lowercase); the ECG is discarded if it contains any of the following keywords:

  `'inversion'`, `'scambio'`, `'errato'`, `'periferic'`, `'elettrod'`, `'sbagliat'`, `'artefatt'`, `'posizionament'`, `'braccia'`.

  This is meant to ensure that the base ECG does not already exhibit positioning anomalies or doubtful reports that are not explicitly encoded among the medical codes.
- **Exclusion through medical codes (quality rejection)**: the JSON contained in the `report` column is parsed. If any of the following rejection codes appears in the diagnostic codes (`codified`), the ECG is ignored:

  `'BTWG01'`, `'BTWG02'`, `'BTWG03'`, `'BTWG04'`, `'BTWG05'`, `'BTWC1109'`, `'BTWC1110'`.

  These are all technical rejection codes (category `rejection`), due for example to duplicate recordings, signals acquired for demonstrative use, non-readable traces, or discrepancies.

- **Sampling and limit**: the IDs surviving all these filters are randomly shuffled (using seed 42 to guarantee reproducibility) and, when necessary, truncated to a desired maximum number. It is possible to set specific data distributions, or to sample examples purely at random, so as to mimic the real clinical distribution. The result is that one can either preserve all possible pathologies with their natural rarity of occurrence, or construct specific distributions to study robustness to a given problem (details will be given later).

3.2.3 Common Preprocessing Pipeline, Quality Assessment, and HDF5 Construction

Regardless of the specific dataset, all ECG traces selected according to the filters of Sec. 3.2.2 are transformed and organized following a common pipeline of signal preprocessing, quality assessment, and construction of the final dataset in HDF5 format.

In summary, for each valid ECG the following steps are applied:

1. **Reading and alignment**: the EDF files compressed in the various ZIP archives are read directly and temporally aligned across all 12 leads, in order to obtain a coherent time $\times$ lead matrix.

2. **Signal preprocessing**: the signals follow a preprocessing pipeline (filtering, noise removal, baseline correction) and are mapped to a standardized form, common to all subsequent analyses and training phases.

3. **Signal Quality Assessment (SQA)**: an algorithmic quality check is applied to the preprocessed signal, both globally and per window. If the global quality of the original signal does not meet the minimum requirements, the entire ECG is discarded; at the same time, local windows that exhibit problems (flatline, excessive noise, major artifacts) are excluded, retaining only those deemed diagnostically reliable.

4. **Normalization (scaling) and windowing**: signals that pass the SQA stage are globally normalized (for example by z-score per lead) and then split into non-overlapping windows of fixed duration (e.g. 2 s at 250 Hz, corresponding to 500 samples). The same mask of valid windows, defined on the preprocessed signal, is applied to the normalized signal, thus preserving coherence between quality control and the data used for training.

5. **Construction of HDF5 datasets and shuffle**: all accepted windows are stored in `.h5` files, divided into training, validation, and test sets, grouping tensors `X` (waveforms) and `Y` (class labels). The HDF5 format is chosen because it is designed for the efficient storage of large multi-channel numerical

time series, such as ECG traces, allowing partial access to data and transparent compression. Finally, tensors are shuffled in RAM blocks to guarantee randomization and fast access during training.

This pipeline constitutes the common basis on which the baseline dataset, the real clinical test set, the `dataset_patologie` and the dataset dedicated to precordial leads are built; the differences between the various datasets mainly concern the clinical selection of patients, the sampling strategies, and the label classes used.

3.2.4 Dataset Creation from the Common Pipeline

Overall, the common steps to be followed for constructing a dataset, starting from the selected ECGs, are the following:

1. **Selection of IDs**: the set of patients/ECGs that will constitute the dataset is identified, applying the technical and clinical filters described in Sec. 3.2.2 and choosing the population of interest (for example "clean" ECGs, patients with axis-deviant pathologies, real traces with electrode inversion, etc.).

2. **Train/test/validation split**: the full set of IDs is divided into training, test, and validation subsets following the proportions 80, 10, 10. The split is performed at patient/ECG level, in order to avoid leakage between sets and to maintain statistical independence between subsets.

3. **Application of the common pipeline**: for each selected ID, the steps of Sec. 3.2.3 are applied (reading EDF files, preprocessing, signal quality assessment, normalization and windowing), obtaining a set of ECG windows associated with their labels, while keeping different recordings separated among subsets.

4. **Synthetic and static inversion (data augmentation)**: clinically "normal" ECGs are taken as a base and synthetically altered versions are produced by mathematically swapping the channels of the leads involved, following the theoretical relationships. Each single normal ECG is simulated in all its possible inversions; this process is performed statically during dataset construction, so as to speed up training and data loading.

5. **Final counting of examples**: in the end, for each of the three subsets (train, validation, test), each ECG yields both the version without inversion and all its variants belonging to the inversion classes considered. In the case of limb leads, the number of ECGs per subset is multiplied by a factor 6 (1 normal class + 5 inversion classes), whereas for precordial leads the factor is equal to 16 (1 normal class + 15 swap classes between two of the six leads).

This number is further multiplied by the number of windows obtained from temporal segmentation (for example 15 windows of 2 s for a 30 s trace), bringing the total number of examples to the order of hundreds of thousands or millions, depending on the dataset.

These steps represent the base pattern for the generation of the datasets described in the following, which reuse the same pipeline while adapting it to different clinical and experimental objectives.

3.2.5 Label Encoding

Before going and further, it is important to clarify the mapping between classes and interchange configurations, used in the dataset creation. Each accepted ECG with class normal was transformed according to the interchange classes defined for the corresponding classification task. The numerical label assigned to each transformed sample was stored in the `Y` dataset of the HDF5 file.

Limb lead transformations

The six input channels of the peripheral lead task were ordered as

$$\mathbf{x}_{\text{limb}} = (I,\, II,\, III,\, aVR,\, aVL,\, aVF)\,. \tag{3.5}$$

In this classification task, six classes were considered: one normal class and five interchange configurations. The mapping between the numerical class encoding and the corresponding transformation is reported in Table 3.2. For the simulations of this task, leads V1-V6 were left unchanged.

Table 3.2: Class encoding and synthetic transformations adopted for the limb-lead classification task. The prime symbol denotes the transformed lead.

Label	**Configuration**	**Transformed limb leads** $(I', II', III', aVR', aVL', aVF')$
0	Normal	$(I,\, II,\, III,\, aVR,\, aVL,\, aVF)$
1	LA–RA	$(-I,\, III,\, II,\, aVL,\, aVR,\, aVF)$
2	RA–LL	$(-III,\, -II,\, -I,\, aVF,\, aVL,\, aVR)$
3	LA–LL	$(II,\, I,\, -III,\, aVR,\, aVF,\, aVL)$
4	Clockwise rotation (`ROT_ORA`)	$(III,\, -I,\, -II,\, aVL,\, aVF,\, aVR)$
5	Counter-clockwise rotation (`ROT_ANT`)	$(-II,\, -III,\, I,\, aVF,\, aVR,\, aVL)$

The classes corresponding to the interchanges involving the neutral electrode RL were not included in the simulation logic, since they are not part of the target

classes of the final limb lead model to be trained. For the unlabelled records, these classes are excluded by the checks of the signal quality, which detects approximately zero leads. The consistency of these two approaches is proved by the fact that when the SQA checks were applied to the real dataset, classes involving the RL were correctly detected and discarded.

Precordial lead transformations

The input channels for the precordial lead task were

$$\mathbf{x}_{\text{prec}} = (V1,\, V2,\, V3,\, V4,\, V5,\, V6)\,. \tag{3.6}$$

Unlike limb electrode transformations, precordial electrodes interchanges were simulated by exchanging the two signals associated with the corresponding anatomical positions. All other precordial channels and all six limb leads were left unchanged.

The numerical class encoding for this task is reported in Table 3.3.

Table 3.3: Class encoding adopted for the precordial-lead classification task.

Label	Configuration	Label	Configuration
0	Normal	8	V2–V5
1	V1–V2	9	V2–V6
2	V1–V3	10	V3–V4
3	V1–V4	11	V3–V5
4	V1–V5	12	V3–V6
5	V1–V6	13	V4–V5
6	V2–V3	14	V4–V6
7	V2–V4	15	V5–V6

3.2.6 Creation of the Baseline Dataset

Once the general construction scheme has been defined, the first dataset created is the baseline dataset, which represents the main reference for training the model for detecting inversions of limb electrodes.

Starting from the “clean” ECGs selected as described in Sec. 3.2.2, approximately 10,000 valid examples without electrode inversion are randomly chosen, so as to mimic the real clinical distribution without imposing artificial constraints on pathology frequencies. All steps of the preprocessing, SQA, normalization, and windowing pipeline described in Sec. 3.2.3 are applied to these IDs, obtaining a set of ECG windows with controlled quality.

Following the scheme in Sec. 3.2.4, data augmentation on limb leads increases the total number of examples from approximately 10k original ECGs to about 60k samples (normal version plus all synthetic inversion variants). Considering the windowing, a 30 s trace is split into 15 non-overlapping windows of 2 s each, bringing the total number of examples available for training close to one million windows.

3.2.7 Real Clinical Test Set for Sim-to-Real Validation

Although the baseline dataset subsets contain testing and validation splits, these are mainly used to monitor training performance on synthetic data. The real challenge is to observe the behaviour of the network on data sampled in real clinical scenarios and to evaluate how far they are from the "mathematical perfection" of inversion simulation.

Unfortunately, real data are not available in sufficient quantity for training; their only useful role is testing for Sim-to-Real validation. By counting all traces that are unsuitable for reporting (`status="rejected"`) and that contain electrode inversion (not belonging to the `"rejection"` category), using the same inversion keywords as before, a total of 1128 recordings is obtained.

On these traces, restricted to limb leads, a manual labelling process was carried out by Eng. Manzini himself, who has extensive experience in ECG signal analysis. The resulting data distribution is the following:

Table 3.4: Distribution of categories in the real limb-lead dataset in order of occurrence.

Category	Count
normal	191
LA-RA	175
ROT_ANTIORARIA	27
LA-LL	14
RA-LL	14
ROT_ORARIA	7
RL-RA	6
RL-LA	1

For a total amount of 435 good examples with duration 30 seconds and all 12 leads regarding limb interchange. In particular, the categories were obtained by remapping the original labels according to the following correspondence:

- *normale* → *normal*
- *RN (Right/Neutre)* → *RL-RA*

- *RL (Right/Left)* → *LA-RA*
- *RF (Right/Foot)* → *RA-LL*
- *LF (Left/Foot)* → *LA-LL*
- *LN (Left/Neutre)* → *RL-LA*
- *antiorario* → *ROT_ANTIORARIA*
- *orario* → *ROT_ORARIA*

All samples with ambiguous label (`"?"`) were excluded.

The selection of traces uses the complement of the IDs from Sec. 3.2.2, because in this case ECGs with electrode inversion are needed, whereas the preprocessing follows the same steps described in Sec. 3.2.3. Another difference lies in the manual labelling of inversion classes and in the fact that this dataset is used exclusively as a real clinical test set for Sim-to-Real validation. As can be clearly seen, the challenge lies in recognizing how rare inversion classes are in general and how some of them are much more frequent than others; for completeness of both analysis and implementation, all of them were retained, even though many works in the literature explicitly discard the rarest classes, such as complete rotations. This represents an additional challenge for our study.

3.2.8 Pathology Dataset for Axis Robustness Stress-Test

To test the model under conditions as close as possible to the real clinical scenario, a dedicated dataset was constructed, named `dataset_patologie`. The main goal is to evaluate the robustness of the model with respect to misalignments of the cardiac electrical axis, which are frequently present in patients and potentially misleading.

Similarly to what was described for the selection of the baseline dataset, also for `dataset_patologie` a preliminary cleaning phase of the clinical database is applied, aimed at guaranteeing the technical and clinical validity of the starting traces, following the common filters of Sec. 3.2.2. In particular, patients with pacemakers are excluded, and all traces suspected of electrode misplacement errors are discarded a priori, identified by the presence in the textual report of keywords such as *"inversion"*, *"scambio"*, *"errato"*, *"sbagliat"*, *"artefatt"*.

To maximize electrical axis variability, the dataset is built by stratified sampling, selecting approximately 20,000 patients divided into four clinical macrogroups (about 5,000 each). The number of patients is double that of the baseline; this choice is not due to scarcity of examples, but to increase variability even further, taking approximately the maximum available occurrences per pathology. The composition is as follows:

- **RAD and LAD (axis deviations)** [∼5,000 traces]: group composed of all patients with right axis deviation (RAD) and left axis deviation (LAD). These conditions produce extremely abnormal electrical axes: RAD in the lower right quadrant (typically between $+90°$ and $+180°$), LAD in the upper left quadrant (between $-30°$ and $-90°$).
- **RBBB (right bundle branch block)** [∼5,000 traces]: introduces delayed activation vectors directed to the right, significantly altering the morphology of the signal and its vector representation.
- **Healthy traces** [∼5,000 traces]: characterized by a physiological electrical axis, typically oriented inferiorly and to the left (standard range between $-30°$ and $+90°$).
- **Mixed pathologies** [∼5,000 traces]: heterogeneous group used as a control, useful to guarantee a higher overall diversity of the dataset.

Also for this dataset, the steps of the common pipeline described in Sec. 3.2.3 are applied, repeating the same stages of preprocessing, SQA, normalization, windowing, and HDF5 storage. The systematic inclusion of pathologies characterized by strong axis deviations (in particular LAD and RAD) forces the model to deal with non-standard cardiac configurations, in which the electrical axis can be significantly rotated, horizontalized, or deviated towards non-physiological quadrants. Consequently, the network cannot simply recognize patterns associated with a "normal" axis, but is forced to learn invariant morphological relationships between leads.

This leads to the development of a more robust and generalizable representation, capable of maintaining high performance even in the presence of significant variations in cardiac axis orientation. The `dataset_patologie` is specifically created as an additional stress-test to verify whether, by training the network on it, significant variations in performance can be observed, given the strong correlation between axis deviation and rotation of the reference system.

3.2.9 Dataset for Precordial Leads

For precordial leads, the analysis is more straightforward, since there is no need for a dedicated real testing dataset; this is due to the fact that precordial leads cannot be reconstructed from one another through linear combinations, since they are not linked by the algebraic relationship of Einthoven's system. From V1 to V6 a subjective point of view is measured, and swapping one precordial electrode with another simply corresponds to swapping their electrocardiographic traces. Consequently, the mathematical simulation of inversion does not need to reflect

any underlying vector reference system, but only to exchange the positions of the two leads.

The examples used are the same as in the baseline dataset, from which only precordial leads are extracted. As for all other datasets, the steps of the common pipeline are applied: technical and clinical filters of Sec. 3.2.2, preprocessing, signal quality assessment, normalization and windowing of Sec. 3.2.3, and finally saving the windows in HDF5 format.

In this case, data augmentation is performed over all the 15 possible swap classes between two of the 6 precordial leads, simulating all exchange configurations among V1, V2, V3, V4, V5 and V6 as:

$$\binom{6}{2} = \frac{6!}{2!(6-2)!} = \frac{6!}{2!4!} = 15. \tag{3.7}$$

This allows training a specific model for the detection of precordial inversions, exploiting the natural independence between these leads and the simplicity of the required mathematical transformation.

3.3 Neural Network Architectures and Training

In this work, lead inversion detection was addressed following the approach proposed by Huang et al., who introduced two separate deep learning models to detect limb lead and chest leads inversion Huang et al. [2024].

The proposed framework follows two classification pipelines. The first is dedicated to limb leads inversion detection and is based on LDenseNet, an architecture designed to extract useful features from ECG windows while keeping the number of trainable parameters limited. The second pipeline is dedicated to precordial leads inversion detection and is based on the Inter Lead Correlation model, which extends the LDenseNet concept. This architecture was designed to address the difficulty of detecting interchanges between chest leads by combining single lead feature extraction with inter lead correlation coefficients Huang et al. [2024].

In this thesis, the architectures originally proposed by Huang et al., in particular the input configuration and the number of output classes, were adopted to the specific use case. For the limb-lead task, the model was trained using the first six ECG channels, corresponding to the limb leads. The limb-lead model was formulated as a six-class classification problem, including one normal class and five abnormal lead configurations: LA-RA reversal, RA-LL reversal, LA-LL reversal, clockwise rotation, and counter-clockwise rotation. For the precordial-lead task, the model was trained using the last six ECG channels, corresponding to leads V1–V6. The precordial-lead model was formulated as a sixteen-class classification problem, including one normal class and fifteen abnormal configurations. These

abnormal classes correspond to all possible pairwise interchanges among the six precordial leads.

The following sections describe the two neural architectures and the training strategy used for model optimization. Finally, the evaluation protocol is presented, including both window-level classification and patient-level aggregation.

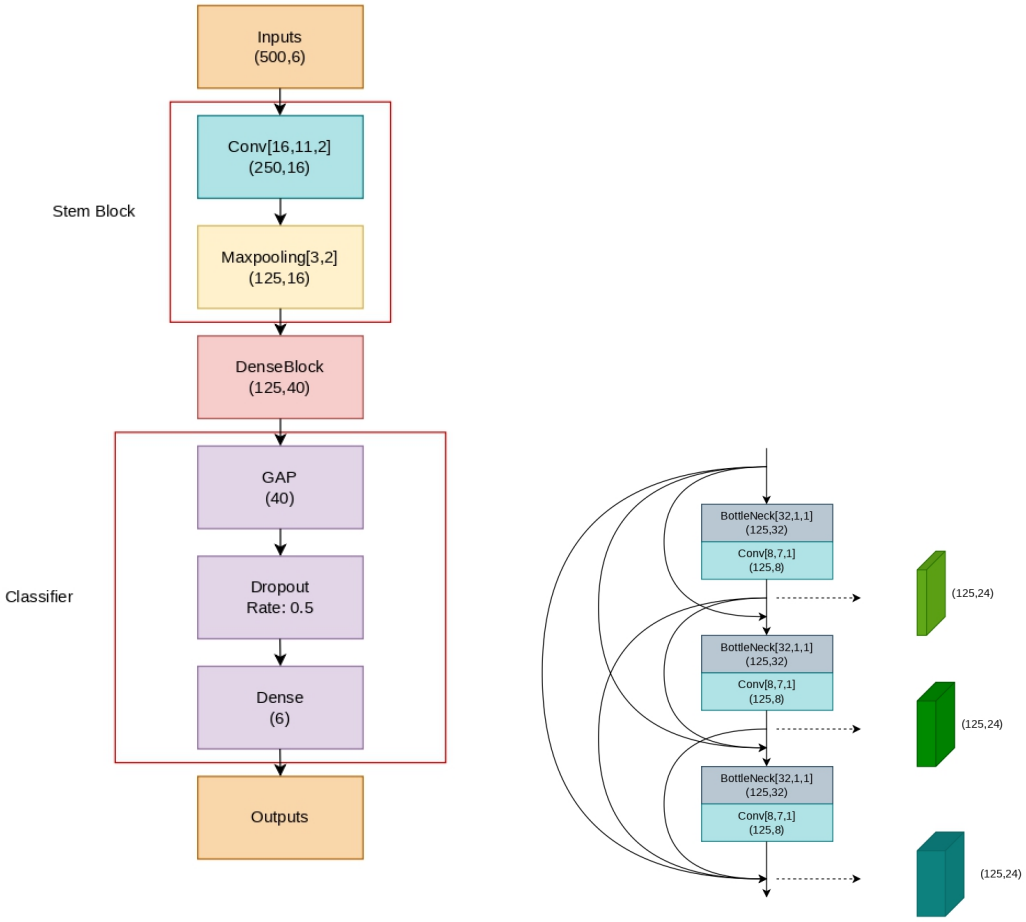
3.3.1 LDenseNet Architecture for Limb-Lead Inversion Detection

For the limb lead inversion detection task, an architecture called LDenseNet was used. It follows the model proposed by Huang et al. Huang et al. [2024], with task specific adaptations, such as the input configuration and the number of output classes. LDenseNet is composed of a Stem Block, a Dense Block and a final classifier.

- **Stem Block.** The Stem Block performs the first level of temporal feature extraction. It consists of a one-dimensional convolution followed by max pooling. The convolution uses 16 filters, a kernel size of 11, stride 2, same padding, and Swish activation. The max-pooling operation further reduces the temporal dimension using a pool size of 3 and stride 2. This initial downsampling allows the model to extract compact representations of the ECG waveform while reducing the computational cost of the following layers.

- **Dense Block.** The output of the Stem Block is then passed to a Dense Block, which is composed of three densely connected convolutional blocks. Each block includes a bottleneck convolution with kernel size 1, followed by a convolution with kernel size 7. The bottleneck layer reduces the computational burden and improves channel-wise interaction, while the second convolution extracts local temporal patterns from the ECG signal. After each block, the output feature maps are concatenated with the input feature maps. This design allows the model to reuse features extracted at different depths and contributes to keeping the architecture compact.

- **Classification Head.** The final module is the classification head, which consists of Global Average Pooling, Dropout and a Dense layer with Softmax activation. Global Average Pooling converts the temporal feature maps into a compact feature vector, while Dropout with rate 0.5 is used for regularization. The final Softmax layer outputs the probabilities for the six limb-lead classes considered.



(a) Overall LDenseNet architecture. (b) Internal structure of the Dense Block.

Figure 3.5: LDenseNet architecture adopted for limb-lead inversion detection. The architecture was adapted from Huang et al. Huang et al. [2024].

3.3.2 Inter Lead Correlation Model for Precordial Leads Interchange Detection

For precordial leads interchange detection, the Inter Lead Correlation (ILC) model was used. It follows the architecture proposed by Huang et al., with adaptations for the sixteen class classification task evaluated in this thesis. The choice of this model is motivated by the fact that, unlike limb leads, which are mathematically related to each other, precordial leads are acquired as separate chest leads placed at different anatomical positions. For this reason, precordial leads interchange just produces an exchange of the signals assigned to two leads positions. Therefore, this task requires a model that evaluates not only the single lead, but also the spatial relationship between all the precordial leads. The Inter Lead Correlation model is designed to capture both the morphology of the single lead and the inter lead relationship that defines the expected anatomic order from V1 to V6 Huang et al. [2024].

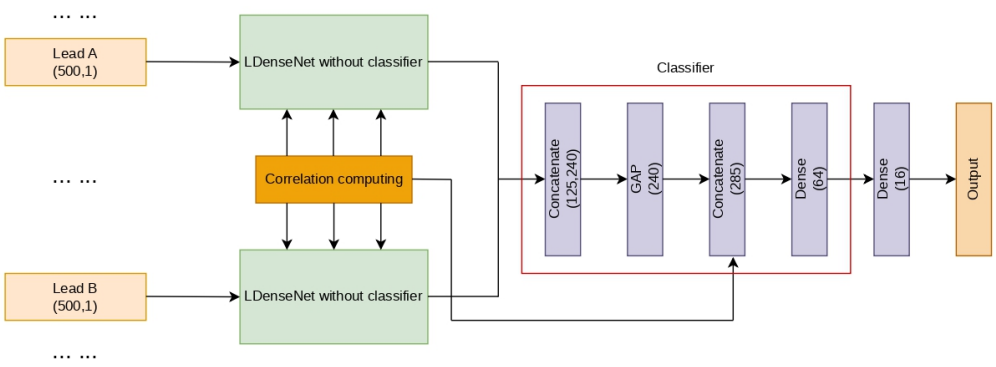


Figure 3.6: ILC architecture adopted for precordial-lead inversion detection. Each precordial lead is processed by an LDenseNet branch with shared weights. The extracted feature maps are combined with the inter-lead correlation coefficients and passed to the final classifier. The architecture was adapted from Huang et al. Huang et al. [2024].

Single lead branches The ILC model is organized as a multi-branch network, in which each branch receives one precordial lead as input and elaborates it with the previous described LDenseNet, which this time does not include a final classifier. So, the branch model consists of a Stem Block followed by a Dense Block. At the end, for each lead, the network returns intermediate feature maps that are later used both for feature aggregation and for inter-lead correlation computation.

An important implementation factor is that the same branch is reused for each lead. This means that the parameters are shared across leads, reducing the number

of trainable parameters.

Inter-lead correlation computation After the single-lead branches have extracted feature maps from each precordial lead, the model computes correlation coefficients between the feature maps of different leads. This is the central component of the ILC architecture. The goal is to provide the classifier with explicit information about how similar or dissimilar the learned representations of different leads are.

For six precordial leads, the number of possible lead pairs is:

$$N_{\text{pairs}} = \binom{6}{2} = 15 \tag{3.8}$$

In the implemented model, correlations are computed at three feature levels, corresponding to the outputs of the three convolutional layers in the Dense Block. Therefore, the total number of correlation coefficients is:

$$N_{\text{cof}} = \binom{6}{2} \times 3 = 45 \tag{3.9}$$

Each coefficient is computed as a normalized dot product between two flattened feature maps:

$$\text{cof}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{m=1}^{M} x_m y_m}{\|\mathbf{x}\|_2 \, \|\mathbf{y}\|_2 + \varepsilon} \tag{3.10}$$

where $\mathbf{x}$ and $\mathbf{y}$ are two feature maps extracted from two different leads, and ε is a small constant added for numerical stability. This operation is equivalent to a cosine-similarity measure between learned feature representations. In this way, the model does not rely only on the morphology of each individual lead, but also on the relationships between pairs of precordial leads Huang et al. [2024].

Since this similarity coefficient is computed using a dot product, it is fair to say that a high positive value indicates similar feature maps, whereas a value close to zero indicates weak similarity and a negative value indicates oppositely oriented representations. This value is not interpreted by the ILC through fixed treshold, but it is concatenated with the convolutional features passed to the classifier, allowing the network to learn whether the observed inter-lead relationships are consistent with the expected spatial organization of the precordial leads.

Feature fusion and classification head The final part of the model is the one destinated to classification. First, the dense feature maps extracted from all six branches are concatenated along the channel dimension. Then a Global Average

Pooling is applied. This step is the one that extracts the morphological features of the individual precordial leads.

Then, the 45 inter-lead correlation coefficients are concatenated with the pooled convolutional features, resulting in a feature vector that contains both information about the individual leads and about the relationship between them.

This combined representation is then passed through a fully connected layer with 64 units and ReLU activation. Finally, a Dense layer with Softmax activation outputs the probability distribution over the sixteen precordial-lead classes. The predicted class is obtained by selecting the output with the highest Softmax probability Huang et al. [2024].

Table 3.5: Main architectural characteristics of the implemented models.

Component	**LDenseNet**	**ILC model**
Input	One tensor with shape $(N, T, 6)$	Six single-lead inputs extracted from a tensor with shape $(N, T, 6)$
Stem Block	16 filters, kernel size 11, stride 2, and Swish activation; max pooling with pool size 3 and stride 2	Per branch: 16 filters, kernel size 11, stride 2, and Swish activation; max pooling with pool size 3 and stride 2
Dense Block	Three units; each unit contains a 1×1 bottleneck convolution with 32 filters, followed by a convolution with 8 filters and kernel size 7	The same three-unit Dense Block is applied independently to each lead through shared branch weights
Final feature maps	40 channels	40 channels per branch and 240 channels after concatenating the outputs of the six branches
Correlation features	Not used	15 lead pairs evaluated at 3 feature levels, producing 45 correlation coefficients
Fused representation	40-dimensional vector after global average pooling	240 pooled morphological features concatenated with 45 correlation coefficients, producing a 285-dimensional feature vector
Classification head	Global average pooling, dropout with rate 0.5, and a 6-unit Softmax layer	Global average pooling, a 64-unit ReLU layer, and a 16-unit Softmax layer
Trainable parameters	9,118	27,336

3.3.3 Training Procedure

In this work, two separate training pipeline were used, one for the limb leads interchange detection and another one for the precordial leads. Even though the two tasks were trained using two different models, the general optimization and training strategy follows the one adopted by Huang et al. for the original LDenseNet and ILC models Huang et al. [2024], with differences in the input dimensions and in the tranining parameters and configurations.

As standard procedure, the model output was interpreted as a probability distribution over the target classes using the last Softmax layer. In both training pipelines, the validation macro F1-score was used as criterion for model selection.

General training strategy

The input format used for training follows the convention expected by TensorFlow convolutional layers:

$$(N, T, C) \tag{3.11}$$

where N is the number of ECG windows, T is the number of temporal samples, and C is the number of input channels.

The training samples were loaded using custom Keras generators, allowing the models to be trained directly from HDF5 files without loading the entire dataset into memory. At the end of each epoch, the training indices were shuffled to reduce the dependency on the order of the samples. The labels were converted to categorical format using one-hot encoding before being returned by the generator.

Both models were trained using categorical cross-entropy as the loss function:

$$\mathcal{L} = -\sum_{c=1}^{C} y_c \log\left(\hat{y}_c\right) \tag{3.12}$$

where C is the number of classes, y_c is the one-hot encoded ground-truth label for class c, and $\hat{y}_c$ is the probability predicted by the Softmax output layer.

The Adam optimizer was used with an initial learning rate of:

$$\eta = 10^{-3} \tag{3.13}$$

During training, accuracy and macro F1-score were monitored. The macro F1-score was used as the main validation metric for saving the best model weights.

Model selection was performed using validation performance. In particular, the best model was defined as:

$$e^* = \arg\max_e F1^{\text{val}}_{\text{macro}}(e), \qquad \theta^* = \theta_{e^*}, \tag{3.14}$$

where e denotes the training epoch and θ_{e^*} denotes the model weights corresponding to the selected epoch.

This selection criterion was preferred over validation accuracy because all classes should contribute equally to the final evaluation, regardless of their frequency or difficulty.

Limb-lead model training

The limb-lead model was trained using the LDenseNet architecture described in the previous section. For this task, the input data were obtained by selecting the first six ECG channels from each window. So, in this case the input is:

$$(N, T, 6) \tag{3.15}$$

The output layer consisted of six neurons, corresponding to the normal class and five limb-lead interchange configurations. The LDenseNet model was trained for a maximum of 50 epochs, using a batch size of 256 and an initial learning rate of 10^{-3}. The model was trained using categorical cross-entropy loss and Adam optimization. Three callbacks were used during training: model checkpointing, early stopping, and learning-rate reduction on plateau. The checkpoint callback saved only the weights of the model with the best validation macro F1-score. Early stopping was applied with a patience of 4 epochs and restored the best weights when training stopped. In addition, the learning rate was reduced by a factor of 0.5 when the validation macro F1-score did not improve for 5 epochs, with a minimum learning rate of 10^{-5}.

After training, the saved weights corresponding to the best validation performance were reloaded and used for the final evaluation.

Hybrid real data and semi-supervised extension for limb-lead training

A large amount of unlabelled clinical ECG data was available in addition to the synthetically generated training set, but could not be directly used in the supervised training pipeline because manual class annotations were missing. To exploit this additional information, a semi-supervised extension of the limb-lead training procedure was implemented. The goal was to use real ECG recordings to complement the simulated training set and improve the model's exposure to real-world signal variability. This was done through a pseudo-labelling strategy, in which a previously trained LDenseNet model was used to assign labels to selected unlabelled ECGs. Only predictions with high confidence were retained, in order to reduce the risk of introducing incorrect labels into the training set.

The semi-supervised approach was organized in four steps. First, candidate ECGs were selected from the database. Only records marked as rejected and

whose text field contained keywords related to the limb leads interchange were selected. Each selected ECG underwent the same preprocessing, normalization, and signal-quality assessment procedure used for the other real clinical recordings.

Second, pseudo-labels were assigned using training weights and parameters from a previously trained LDenseNet model. For each ECG, the model produced a probability vector for every valid window. These window-level probabilities were then averaged to obtain an ECG-level/patient-level prediction:

$$\bar{\mathbf{p}} = \frac{1}{W} \sum_{w=1}^{W} \mathbf{p}_w \tag{3.16}$$

where W is the number of valid windows and $\mathbf{p}_w$ is the Softmax probability vector predicted for window w.

The pseudo-label was assigned as:

$$\hat{y} = \arg \max_{c \in \{1,\ldots,C\}} \bar{p}_c \tag{3.17}$$

where $C = 6$ is the number of limb-lead classes and $\bar{p}_c$ is the averaged probability associated with class c.

The corresponding confidence score was defined as:

$$s = \max_{c \in \{1,\ldots,C\}} \bar{p}_c \tag{3.18}$$

Only ECGs with confidence greater than or equal to 0.95 were accepted as pseudo-labelled samples:

$$s \geq 0.95 \tag{3.19}$$

This threshold was used to ensure that only the records of which the network was sure were added to the training set.

In addition to the pseudo labelled recordings, manually annotated real ECGs that had not been assigned to the independent clinical test set were added to the real data training pool. These recordings had their clinical ground truth labels, so they were not treated as pseudo labelled samples and they were separated from the records included in the test set by performing a check at record level, preventing the same clinical ECG from contributing to both training and final evaluation.

Third, a new hybrid training set was constructed, adding to the original simulated training set the new pseudo labelled records and the manually labelled real ECGs. The pseudo-labelled windows were upsampled in order to represent approximately 10% of the final training set, with a maximum replication factor of 15. During the upsampling stage, data augmentation was applied to the replicated pseudo-labelled windows. This step was introduced to avoid adding multiple identical copies of the same ECG windows to the training set, which could increase

the risk of overfitting. In particular, Gaussian noise with zero mean and standard deviation 0.03 and random amplitude scaling were applied to the replicated windows in order to simulate small acquisition variability while preserving the assigned pseudo-label. The combined dataset was shuffled and written to a new HDF5 file.

Finally, LDenseNet was retrained from scratch using the newly constructed hybrid training set. The validation set remained the original simulated validation set, while the real labelled test set was kept unchanged for final evaluation. This ensured that the semi-supervised data were used only during training and did not affect the independent test protocol.

The training configuration was kept consistent with the original limb-lead training setup: the model was trained for a maximum of 50 epochs, using a batch size of 256, an initial learning rate of 10^{-3}, categorical cross-entropy loss, and Adam optimization. Model checkpointing, early stopping, and learning-rate reduction on plateau were again based on the validation macro F1-score.

Overall, this hybrid extension was designed to expose the limb-lead classifier to a larger number of real ECG morphologies. The manually labelled recordings provided reliable real examples, while the high confidence pseudo labelled ECGs made it possible to exploit additional unlabelled records.

Precordial-lead model training

The precordial-lead model was trained using the Inter-Lead Correlation architecture. For this task, the input data were obtained by selecting the last six ECG channels, corresponding to the precordial leads V1–V6. As in the limb-lead pipeline, input tensors have shape:

$$(N, T, 6) \tag{3.20}$$

The output layer consisted of sixteen neurons, corresponding to one normal class and fifteen precordial-lead interchange configurations.

The ILC model was trained for a maximum of 20 epochs, using a batch size of 256 and an initial learning rate of 10^{-3}. As for the limb-lead model, categorical cross-entropy was used as the loss function and Adam was used as the optimizer. Model checkpointing and early stopping were applied by monitoring the validation macro F1-score. The checkpoint callback saved the best model weights, while early stopping was applied with a patience of 12 epochs and restored the best weights at the end of training.

A longer early-stopping patience was used for the precordial model because the ILC architecture is more complex than LDenseNet and the classification task includes a larger number of classes. This allowed the model more training epochs to improve validation performance before stopping.

Table 3.6: Training configurations adopted for the three experiments.

Setting	**Limb supervised**	**Limb semi-supervised**	**Precordial**
Model	LDenseNet	LDenseNet	ILC
Maximum epochs	50	50	20
Batch size	256	256	256
Initial learning rate	10^{-3}	10^{-3}	10^{-3}
Optimizer	Adam	Adam	Adam
Loss function	Categorical cross-entropy	Categorical cross-entropy	Categorical cross-entropy
Selection metric	Validation macro F1-score	Validation macro F1-score	Validation macro F1-score
Early-stopping patience	4 epochs	4 epochs	12 epochs
Learning-rate reduction	Factor 0.5, patience 5, minimum 10^{-5}	Factor 0.5, patience 5, minimum 10^{-5}	Not applied
Pseudo-label threshold	Not applicable	0.95	Not applicable
Approximate pseudo-labelled share	Not applicable	10% of the final training set	Not applicable
Pseudo-labelled augmentation	Not applicable	Gaussian noise and random amplitude scaling	Not applicable
Initialization	From scratch	From scratch	From scratch

3.3.4 Evaluation Protocol

After training, the evaluation step was performed by reloading the weights corresponding to the best validation F1-score. The evaluation was performed separately for the limb-lead model and for the precordial-lead model, but it follows the same method: the predicted class is assigned selecting the class with the highest probability after the Softmax layer:

$$\hat{y}_i = \arg\max_c p_{i,c} \tag{3.21}$$

where $p_{i,c}$ is the probability assigned by the model to class c for window i.

These predicted labels were compared to the ground truth labels and a confusion matrix was built for a better understanding of the model performance. The detailed definition of the metrics and the resulting performance values are reported in the following chapter.

The evaluation was first performed at window level, and then these results were used to compute the patient/ECG-level classification. The model first produces a Softmax probability vector for each ECG window. Then all window belonging to the same patient ECG were grouped to obtain the patient-level evaluation. For each patient, the windows probability vector were averaged:

$$\bar{\mathbf{p}}_r = \frac{1}{W_r} \sum_{w=1}^{W_r} \mathbf{p}_{r,w} \tag{3.22}$$

where W_r is the number of windows associated with patient r, and $\mathbf{p}_{r,w}$ is the probability vector predicted for window w of that patient.

Then, the label was assigned by choosing the class with the highest probability:

$$\hat{y}_r = \arg\max_c \bar{p}_{r,c} \tag{3.23}$$

Normal-versus-abnormal evaluation

During the evaluation step, in addition to the multi-class evaluation, a binary classification task was performed. The goal was to analyse how much the model is able to distinguish between normal ECGs and ECGs with an inversion, indipendently from the particular inversion class. This task is particularly relevant in the context of this work beacause the two types of errors do not have the same impact: if an abnormal ECG is incorrectly classified as normal is considered more critical than a normal ECG being classified as abnormal. Indeed, the latter may just lead to an unnecessary additional check, while the former may allow an incorrectly acquired ECG to remain undetected, together with all the potential consequences for ECG interpretation and subsequent clinical decisions.

No separate model was trained for this binary task. The same models trained for multi-class classification were used, and the binary analysis was performed as a post-processing step during evaluation. After obtaining the multi-class predictions, both the ground-truth labels and the predicted labels were mapped into two categories: class 0 for normal ECGs and class 1 for ECGs affected by any type of lead inversion:

$$y^{\text{bin}} = \begin{cases} 0, & \text{if } y = 0, \\ 1, & \text{if } y > 0. \end{cases} \tag{3.24}$$

For the limb-lead task, all classes from 1 to 5 were grouped into the abnormal class. For the precordial-lead task, all classes from 1 to 15 were grouped into the abnormal class.

The metrics used to quantify the performance of the multi-class and binary evaluations are described in the next chapter, together with the experimental results.

3.4 The Geometric Limit of Limb Leads and V6

Accidental inversion of limb cables during an instrumental electrocardiographic examination is a well-known source of artifacts and misinterpretation in clinical practice Batchvarov et al. [2007, 2023]. It does not introduce random distortion into the signal, but it produces a deterministic transformation of its representation in the frontal plane; the six peripheral leads are not independent measurements, they are linear projections of cardiac electrical activity onto Einthoven's reference system, which is defined in the frontal plane and is commonly used to estimate the mean electrical axis of the heart Macfarlane et al. [2010], Kuhn and Rose [2008]. As a result, misplacement of the limb electrodes can be described in terms of geometric operators that modify the apparent orientation of the axis, rather than arbitrary perturbations.

This geometric viewpoint has an important consequence: the native electrical axis of the patient becomes a structural confounding factor. Depending on its angular position, the same technical error may appear very obvious or almost completely invisible. In contrast, different errors applied to patients with different native axes can generate peripheral tracings that are indistinguishable. The problem of classifying limb lead inversions is therefore not only a question of pattern recognition but also a problem of identifiability of the information that remains observable in the frontal projection.

3.4.1 Global Table of Angular Transformations

To visualize the behaviour of the different inversion patterns, it is useful to report the complete map of all the possible angular transformations on a set of standard cardiac axes in the frontal plane. The following table shows, for each class, the native axis α and the corresponding apparent axis α'. In this way, one can directly observe how each technical error acts as a geometric operator on the cardiac vector and how heterogeneous transformations may converge to the same apparent values.

Table 3.7: Transformation of standard cardiac axes for the different limb-lead inversion classes.

Class	Inversion	$\alpha = -30°$ α'	$\alpha = 0°$ α'	$\alpha = +30°$ α'	$\alpha = +60°$ α'	$\alpha = +90°$ α'
0	Normal	$-30°$	$0°$	$+30°$	$+60°$	$+90°$
1	LA–RA	$-150°$	$\pm 180°$	$+150°$	$+120°$	$+90°$
2	RA–LL	$-30°$	$-60°$	$-90°$	$-120°$	$-150°$
3	LA–LL	$+90°$	$+60°$	$+30°$	$0°$	$-30°$
4	ROT_ORA	$+90°$	$+120°$	$+150°$	$\pm 180°$	$-150°$
5	ROT_ANT	$-150°$	$-120°$	$-90°$	$-60°$	$-30°$

Recall that a physiological axis is in between -30° and 90° with a maximal deviation in the range of -90° and 180°. For completeness, the same analysis can be extended to physiologically extreme axes that are often used in theoretical discretizations, such as +120°. The behaviour remains consistent with the same transformation rules:

Table 3.8: Extension of angular transformations to the axis $\alpha = +120°$.

Class	Inversion	$\alpha' = f(\alpha)$ for $\alpha = +120°$
0	Normal	$+120°$
1	LA–RA	$+60°$
2	RA–LL	$\pm 180°$
3	LA–LL	$-60°$
4	ROT_ANT	$0°$
5	ROT_ORA	$-120°$

A joint inspection of Tables 3.7 and 3.8 shows that different transformations can lead to the same apparent axis even when starting from different native axes. This observation is the basis for the collapse families, that is, the configurations in which the system loses the ability to uniquely discriminate the origin of the tracing.

3.4.2 Vector Mimicry and Information Collapse

The central phenomenon emerging from the analysis is *vector mimicry*. This term indicates the possibility that two physiologically different configurations, that is, two distinct pairs composed of the native cardiac axis and the positioning error, are able to produce the same apparent mean axis in the frontal plane.

Under this simplified two dimensional model, the corresponding limb lead projections may therefore coincide. This does not necessarily mean that the complete ECG records are identical, since they also contain time dependent morphology and three dimensional information not included in the simplified mean axis model.

Formally, vector mimicry arises because the observed system is only a partial projection of the electrical state of the heart. When two different transformations act on appropriately chosen native vectors, they can converge to the same apparent vector. The result is a collapse of the original clinical information: what was distinct in the physiological space becomes indistinguishable in the observable space. This mechanism helps to explain why some inversion patterns are classified with high precision whereas others remain challenging and require dedicated signal quality and misplacement detection methods Rjoob et al. [2020], Jekova et al. [2016], Huang et al. [2024], Rafiei et al. [2026].

3.4.3 Mathematical Model of the Transformations

Let $\mathbf{v}(\alpha)$ be the vector that represents the electrical activity of the heart in the frontal plane for a patient with a native axis α. Each limb lead configuration can be modelled by a linear transformation T_k, where the index k identifies a specific class among Normal, LA–RA, RA–LL, LA–LL, ROT_ANT and ROT_ORA. The apparent axis observed after the transformation can be written as:

$$\mathbf{v}'_k = T_k \mathbf{v}(\alpha). \tag{3.25}$$

In angular terms, one can introduce a projection function

$$\alpha'_k = f_k(\alpha), \tag{3.26}$$

where f_k maps the native axis α to the apparent axis α'_k produced by the transformation T_k. The entire problem of mimicry can thus be studied by comparing these projection functions.

From a conceptual point of view, the key step is to recognize that classification performance does not depend only on the morphology of the waveform, but also on the structure of the map f_k itself. If two different maps take the same value for appropriately chosen native axes, then the observable system loses the ability to distinguish between the two underlying physical causes.

3.4.4 General Law of Mimicry

The first form of collapse involves different patients, that is, different native axes. Let T_A and T_B be two transformations associated with two distinct technical errors, and let f_A and f_B be the corresponding angular functions. The condition for mimicry can be written as:

$$f_A(\alpha) = f_B(\beta), \qquad \alpha \neq \beta. \tag{3.27}$$

This relationship defines the **General Law**: two patients with different native axes α and β, subjected to two different inversion patterns, collapse to the same apparent axis α'. Under these conditions, the mean frontal axis are identical, even though this does not mean that the ECG records will be identical, since there could be differences in waveform morphology, temporal evolution, three-dimensional cardiac activity, and patient anatomy.

A particularly illustrative example is illustrated in 3.7 by Family 2, which relates the LA–RA inversion to the anterior rotation ROT_ANT. In this case, the collapse law can be expressed as:

$$\alpha = +30^\circ \quad \text{for LA–RA}, \qquad \beta = -90^\circ \quad \text{for ROT_ANT} \tag{3.28}$$

with a common apparent axis

$$\alpha' = +150^\circ. \tag{3.29}$$

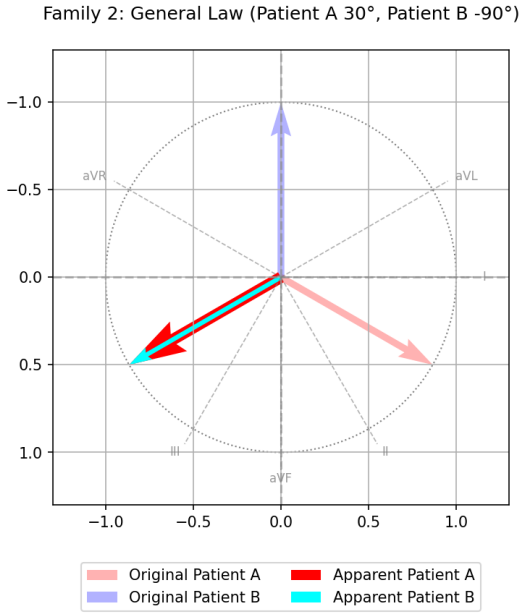
This means that a patient with native axis $+30^\circ$ affected by LA–RA inversion and a patient with native axis -90° affected by ROT_ANT produce the same apparent mean frontal axis under the adopted model. From a diagnostic perspective, the physiological difference between the two patients is completely absorbed by the geometry of the transformation, making the configurations ambiguous.

An analogous behaviour can be observed in Family 1 in figure 3.8, where a patient A with native axis $+60^\circ$ and LA–LL inversion is compared with a patient B with native axis 0° in normal configuration: in this case as well, the transformations lead to a common apparent axis under the simplified model.

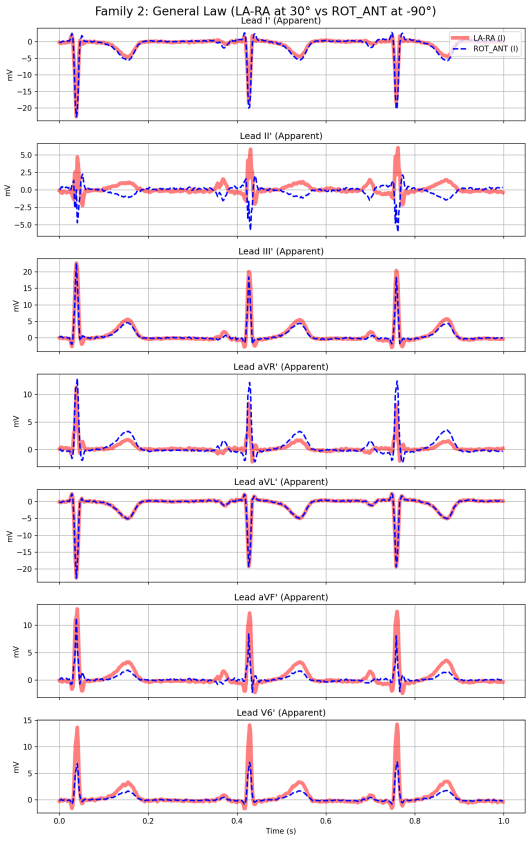
3.4.5 Eigen-Space and Null Space

The second form of collapse is even deeper, because it concerns the same patient, or also, two different patients with the same native axis. In this case, the question is whether two different transformations can produce the same output starting from the same input vector. The condition can be written as:

$$T_A\mathbf{v}(\alpha) = T_B\mathbf{v}(\alpha). \tag{3.30}$$

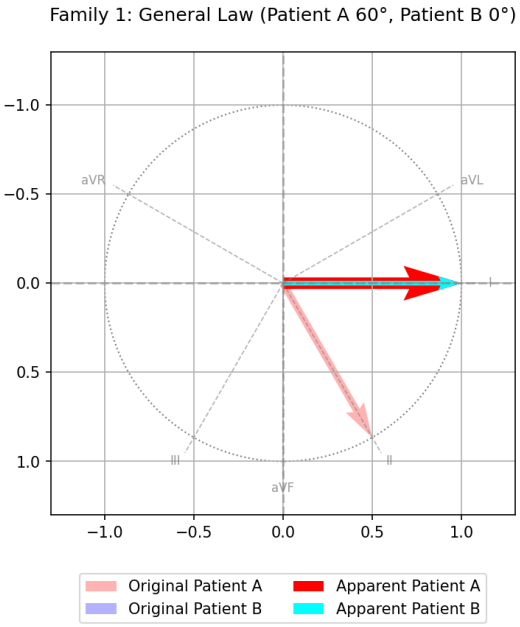


Polar diagram of the vectors.

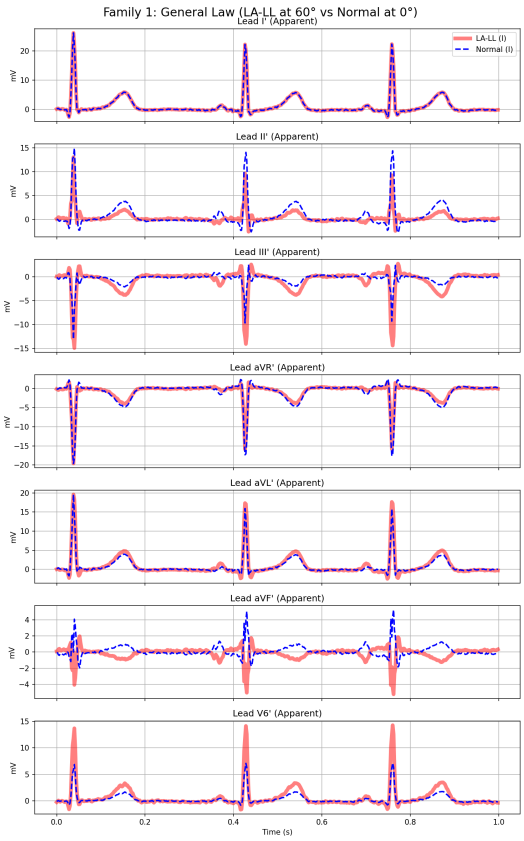


Apparent tracings of I, II, III, aVR, aVL, aVF and V6.

Figure 3.7: Family 2 under the General Law (LA–RA vs ROT_ANT).



Polar diagram for Family 1 (LA–LL at 60° vs Normal at 0°).



Apparent tracings of I, II, III, aVR, aVL, aVF and V6 for Family 1 (LA–LL vs Normal) under the General Law.

Figure 3.8: Family 1 under the General Law: polar diagram and apparent frontal tracings.

Moving all terms to the left-hand side yields

$$(T_A - T_B)\mathbf{v}(\alpha) = \mathbf{0}. \tag{3.31}$$

Since the cardiac electrical activity is never the zero vector, the equation admits physically meaningful solutions only if the difference matrix is singular, that is, if

$$\det(T_A - T_B) = 0. \tag{3.32}$$

The set of axes α that satisfy this condition belong to the **Null Space** of the pair of transformations. From a geometric viewpoint, this Null Space can be interpreted as the **Eigen-space** of the problem: an angular region in which two distinct technical errors become indistinguishable even when applied to the same patient. In this scenario, there is not merely a strong similarity between different configurations, but a complete coincidence of the observable information.

The structure of this Eigen-space can be summarized in table 3.9, which lists all perfect collapse points for the main pairs of limb lead transformations.

For Family 2 (LA–RA vs ROT_ANT) in figure 3.9, the Eigen-space is identified by the point $\alpha = -30°$: at this coordinate, both transformations project the cardiac vector onto the apparent axis $\alpha' = -150°$, resulting in coinciding idealized limb lead projections in the simplified model.

Similarly, in Family 1 (Normal vs LA–LL) in figure 3.10, the point $\alpha = +30°$ corresponds to a condition of geometric self-cancellation: the LA–LL inversion is completely absorbed by the structure of the native vector and the apparent axis coincides with the physiological one.

For Family 3 (RA–LL vs ROT_ORA) in figure 3.11, the Eigen-space lies at $\alpha = -90°$, with a common apparent axis $\alpha' = +30°$. Here, also, the two inversions become physically indistinguishable in the frontal plane and in the limb leads.

3.4.6 Analytical Interpretation of the Role of V6

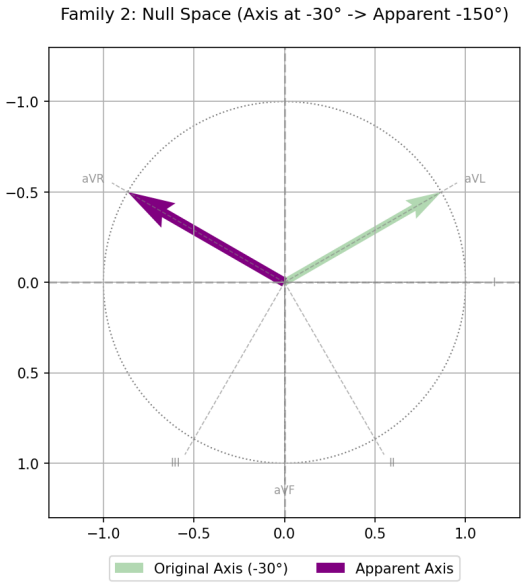
The theory of geometric collapse clarifies the contribution of the chest lead **V6**. Limb leads observe the heart exclusively in the frontal plane; therefore, any ambiguity generated within that plane cannot be resolved by staying confined to the same observable space made by six linearly dependent axes. Adding V6 changes this scenario because it introduces additional information that is sensitive to the three-dimensional component of electrical propagation.

Outside the Null Space, two patients that collapse to the same apparent frontal axis still retain a residual morphological difference along the direction observed by V6. This difference is sufficient for the neural network to separate the two classes. Mathematically, V6 can be interpreted as an additional function $g(\alpha)$ that extends the observation space:

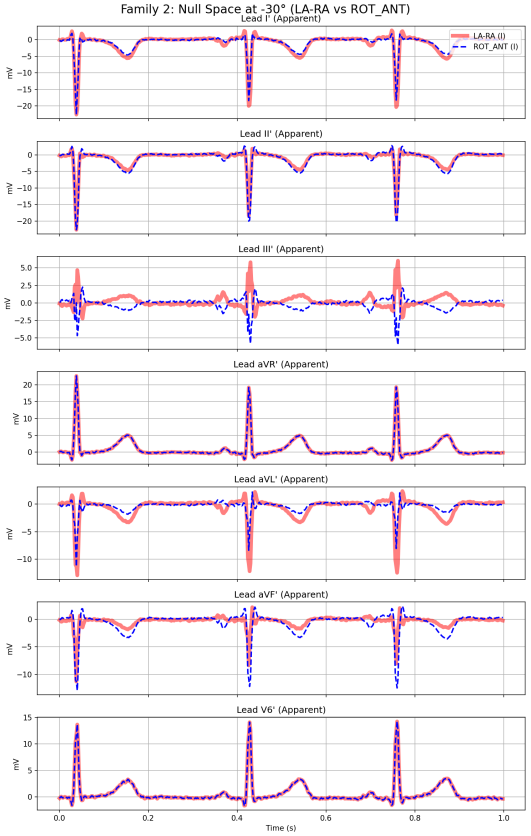
$$\mathcal{O}_{6L} = \{I, II, III, aVR, aVL, aVF\}, \qquad \mathcal{O}_{7L} = \mathcal{O}_{6L} \cup \{V6\}. \tag{3.33}$$

Table 3.9: Nine points of perfect collapse (Null Space) between pairs of limb-lead transformations.

Pair of transformations	Native axis α	Apparent axis α'	Interpretation
Normal vs LA–LL	$+30°$	$+30°$	Geometric self-cancellation of the LA–LL error.
Normal vs RA–LL	$-30°$	$-30°$	RA–LL same transformed frontal vector as the normal configuration in the simplified model.
Normal vs LA–RA	$+90°$	$+90°$	Collapse on an invariant vertical axis.
LA–RA vs ROT_ANT	$-30°$	$-150°$	Two distinct errors produce the same extreme apparent axis.
LA–RA vs ROT_ORA	$+30°$	$+150°$	Collapse on the opposite extreme axis.
RA–LL vs ROT_ANT	$+30°$	$-90°$	Convergence on an inferior-left apparent axis.
RA–LL vs ROT_ORA	$-90°$	$+30°$	Eigen-space of Family 3, located at $-90°$.
LA–LL vs ROT_ANT	$+90°$	$-30°$	Collapse on an oblique inferior-right axis.
LA–LL vs ROT_ORA	$-30°$	$+90°$	Collapse on a superior vertical axis.

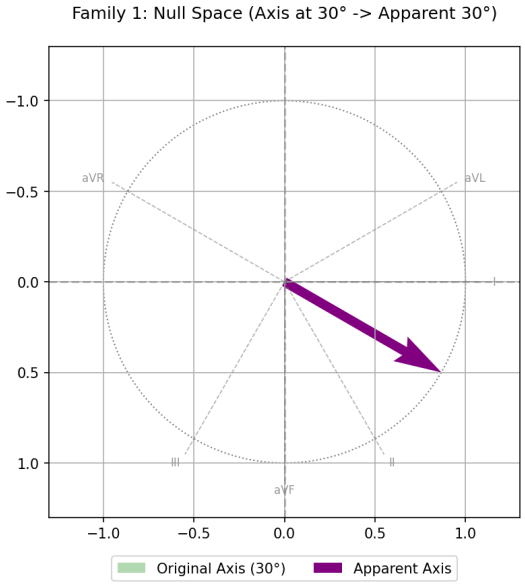


Polar representation of the Eigen-space for Family 2 (native axis $-30°$, apparent axis $-150°$, LA–RA vs ROT_ANT configuration).

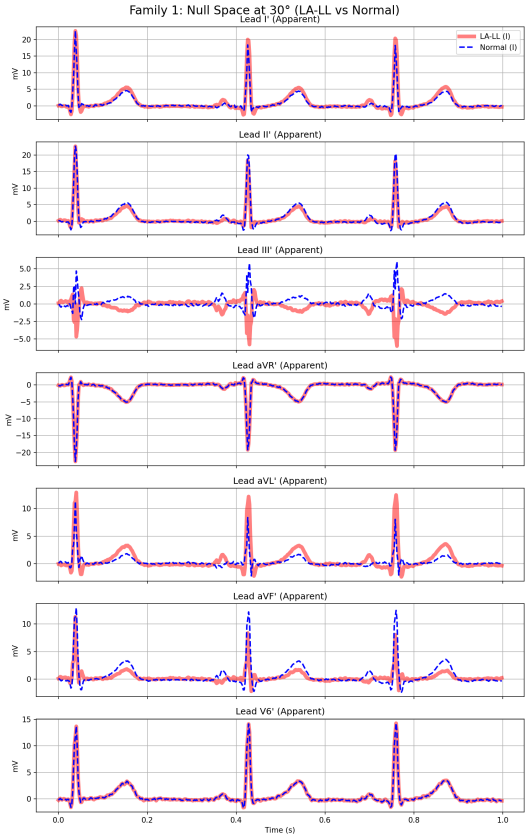


Apparent tracings of I, II, III, aVR, aVL, aVF and V6 for Family 2 in Eigen-space ($\alpha = -30°$).

Figure 3.9: Family 2 in Eigen-space: polar representation and apparent frontal tracings.

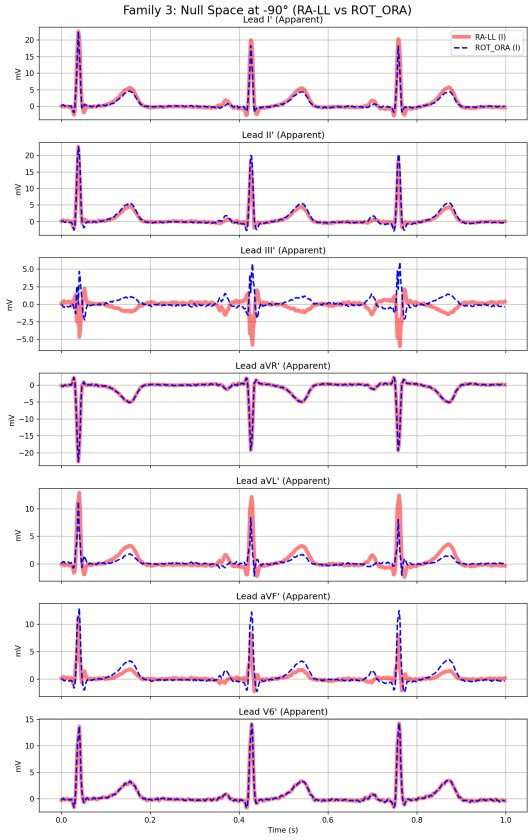


Polar diagram of the Eigen-space for Family 1 (native axis +30°, apparent axis +30°, Normal vs LA–LL configuration).

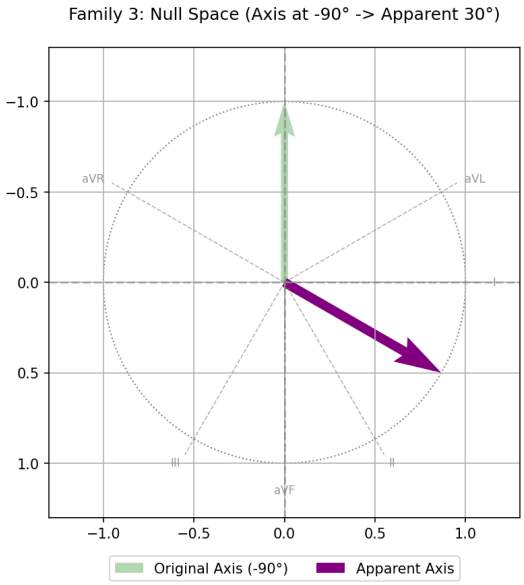


Apparent tracings of I, II, III, aVR, aVL, aVF and V6 for Family 1 in Eigen-space ($\alpha = +30°$).

Figure 3.10: Family 1 in Eigen-space: polar diagram and apparent frontal tracings.



Polar diagram of the Eigen-space for Family 3 (native axis $-90°$, apparent axis $+30°$, RA–LL vs ROT_ORA configuration).



Apparent tracings of I, II, III, aVR, aVL, aVF and V6 for Family 3 in Eigen-space ($\alpha = -90°$).

Figure 3.11: Family 3 in Eigen-space: polar diagram and apparent frontal tracings.

When two configurations satisfy

$$f_A(\alpha) = f_B(\beta), \qquad \alpha \neq \beta, \tag{3.34}$$

they are indistinguishable in the system $\mathcal{O}_{6L}$, but may become separable in the system $\mathcal{O}_{7L}$ if

$$g_A(\alpha) \neq g_B(\beta). \tag{3.35}$$

This is why V6 resolves the **General Law**: collapse in the frontal plane does not necessarily imply collapse in the full three-dimensional signal space.

3.4.7 The Absolute Limit: when V6 is not enough

The situation changes radically inside the **Null Space**. In this case, it is not only the frontal projection of two different configurations that collapses, but the entire useful spatial representation. When the patient lies at an Eigen-space coordinate, the two transformations produce a complete coincidence of the observable output; consequently, the additional information provided by V6 can also become identical.

Formally, there exists a critical axis $\alpha^\star$ such that

$$T_A\mathbf{v}(\alpha^\star) = T_B\mathbf{v}(\alpha^\star), \tag{3.36}$$

and, at the same time, all observable components are superimposed. In this condition, the problem is no longer classificatory but ontological: the information that would allow for the distinction between the two cases is simply not present in the data. No model, regardless of its complexity, can recover information that the physics of projection has completely removed.

This observation is fundamental because it precisely defines the meaning of the model's performance. Errors concentrated near Eigen-space points do not represent a failure of the algorithm, but the necessary manifestation of a structural limit of incomplete limb-lead electrocardiography in particular scenarios.

3.4.8 Clinical and Theoretical Interpretation

The presence of geometric collapse regions changes how the classifier's performance should be interpreted. A network error does not always have the same meaning: in some cases it reflects statistical limits or noise, but in others it expresses the inevitable effect of a mathematical degeneration of the signal. This distinction makes it possible to separate stochastic errors, due to tracing artifacts or suboptimal signal quality, from errors related to electrode misplacement and lead configuration Satija et al. [2018], Xie et al. [2021], Rjoob et al. [2020].

Within this framework, adding **V6** does not simply improve the network in a generic way, but expands the observation space in a targeted fashion, recovering the

three-dimensional information needed to break the ambiguities of the General Law. The General Law describes situations in which different configurations become indistinguishable in the frontal plane yet remain separable thanks to additional chest information such as V6. In other words, collapse in the frontal projection does not necessarily imply collapse in the full three-dimensional signal space.

The Eigen-space, by contrast, marks the extreme limit of the system. In Null Space regions, projection permanently erases the differential information between two configurations: ambiguity is not due to a weakness of the model, but to the physical loss of information. In these conditions, the problem is no longer classificatory but one of observability; the data simply do not contain enough information to distinguish the possible causes.

From this perspective, electrocardiography based solely on limb leads is intrinsically limited by the geometry of Einthoven's plane, and the problem of electrode inversions must be understood as a problem of partial observability of a vector system. The contribution of **V6** should therefore be interpreted not as a simple performance boost, but as a geometric extension of the observable space. The neural network is effective because it exploits all available information outside the collapse regions; inside the Null Space, by contrast, the system reaches its event horizon: beyond this threshold, clinical distinction between certain inversion patterns ceases to be observable.

Chapter 4

Experimental Results

This chapter reports the results obtained with the proposed models for automatic recognition of ECG electrode interchange. The analysis is organized by lead group: limb leads and precordial leads. For limb leads, we further distinguish between a fully simulated scenario (simulated-to-simulated) and a transfer scenario to real clinical signals (simulated-to-real).

For each scenario, we report both multi-class performance, where each interchange configuration is treated as a separate class, and binary performance, where all abnormal conditions are collapsed into a single interchange class and compared with the normal configuration. The binary view is useful for assessing the model's ability to detect whether an electrode interchange is present at all, independently of its specific class.

The results are reported at window and ECG level. The interpretation, clinical implications and methodological limitations are discussed in Chapter 5.

4.1 Evaluation Metrics

The model performance was evaluated using standard metrics: accuracy, precision, sensitivity, specificity, F1-score, AUROC and AuPRC Altman [1994], Pepe [2003], Author [2023], Šimundić [2010], Hajian-Tilaki [2013].

For the multi-class evaluation, precision, sensitivity (recall), specificity and F1-score were computed independently for each class. For a given class c, samples belonging to c were considered positive, whereas samples belonging to all other classes were considered negative. The corresponding counts are denoted by TP_c, TN_c, FP_c, and FN_c.

In addition to class-wise metrics, overall multi-class accuracy was computed from the complete confusion matrix.

Sensitivity (recall) Sensitivity measures the proportion of samples belonging to that class that are correctly identified:

$$\text{Sensitivity} = \frac{TP}{TP + FN}. \tag{4.1}$$

In clinical applications, this is a key metric, because a low number of false negatives reduces the risk of missing truly abnormal tracings Author [2023].

Specificity For a given class c, specificity measures the proportion of samples not belonging to that class that are correctly identified as negative:

$$\text{Specificity} = \frac{TN}{TN + FP}. \tag{4.2}$$

High specificity corresponds to few false alarms and a low probability of labelling as abnormal patients with correct electrode placement Altman [1994], Author [2023].

Overall accuracy For the multi-class confusion matrix, the overall performance was evaluated with accuracy.

Given a confusion matrix $\mathbf{C}$ with K classes, it is defined as:

$$\text{Accuracy} = \frac{\sum_{c=1}^{K} C_{cc}}{N}, \tag{4.3}$$

where C_{cc} denotes the number of correctly classified samples of class c, and N is the total number of evaluated samples.

For the binary evaluation, accuracy is defined as:

$$\text{Accuracy} = \frac{TP + TN}{N}. \tag{4.4}$$

where N is TP + TN + FP + FN.

Accuracy provides a global view of model performance, but in the presence of a class imbalance it can be overly optimistic and is therefore interpreted together with sensitivity and specificity Altman [1994], Šimundić [2010].

Precision and F1-score Precision measures the proportion of correctly predicted positives among all predicted positives, so the proportion of samples correctly assigned to a class among all samples predicted as belonging to it:

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c}. \tag{4.5}$$

The F1-score combines precision and sensitivity into a single index through their harmonic mean:

$$\mathrm{F1}_c = 2 \cdot \frac{\mathrm{Precision}_c \cdot \mathrm{Sensitivity}_c}{\mathrm{Precision}_c + \mathrm{Sensitivity}_c}. \tag{4.6}$$

It is widely used in machine learning because it penalizes scenarios in which one of the two components is much lower than the other, providing a compact indication of the trade-off between false positives and false negatives Šimundić [2010].

AUROC and AuPRC The area under the ROC curve (AUROC) summarizes the discriminative ability of the model as the decision threshold varies: each point on the curve represents a sensitivity–specificity pair, and the overall area measures how much the model tends to assign higher scores to subjects with misplacement than to normal subjects Hajian-Tilaki [2013], Pepe [2003]. The area under the Precision–Recall curve (AuPRC) instead focuses on precision and sensitivity for the positive class, and is particularly useful when the prevalence of misplacement is low Pepe [2003], Hajian-Tilaki [2013].

4.1.1 Limb Leads: Simulated-to-Simulated Scenario

Window-level results

In the simulated test set for limb leads, the baseline supervised model reaches an overall multi-class accuracy of 96.61%, with AUROC equal to 0.9969 and AuPRC equal to 0.9886.

Per-class metrics at window level are reported in Table 4.1.

Table 4.1: Per-class performance on limb leads (simulated-to-simulated scenario, window level).

Class	Acc.	Prec.	Rec.	Spec.	F1
normal	0.9891	0.9743	0.9596	0.9949	0.9669
LA–RA	0.9892	0.9657	0.9698	0.9931	0.9678
RA–LL	0.9880	0.9781	0.9492	0.9957	0.9634
LA–LL	0.9885	0.9560	0.9758	0.9910	0.9658
ROT_ORA	0.9887	0.9551	0.9780	0.9908	0.9664
ROT_ANT	0.9887	0.9681	0.9640	0.9937	0.9661

All classes show accuracy values above 0.988 and F1-scores between 0.9634 and 0.9678.

Binary analysis: normal versus abnormal Collapsing all misplacement classes into a single "abnormal" category, the model reaches a binary accuracy of 0.9891. Specificity on normal tracings is 0.9596, while sensitivity towards abnormal tracings is 0.9949; binary precision is 0.9919 and the F1-score is 0.9934. The corresponding confusion matrix in figure 4.1 reports 60 975 abnormal windows correctly identified and 11 762 normal windows correctly classified, with 310 false negatives and 495 false positives.

Patient-level results

In order to obtain an ECG-level prediction, the window-level probability vectors belonging to the same ECG were averaged, and the final prediction was the class with the highest mean probability.

The overall accuracy increased from 96.61% at window level to 97.53% at ECG level, with AUROC equal to 0.9969 and AuPRC increased from 0.9886 to 0.9889. Per-class metrics at patient level are summarized in Table 4.2.

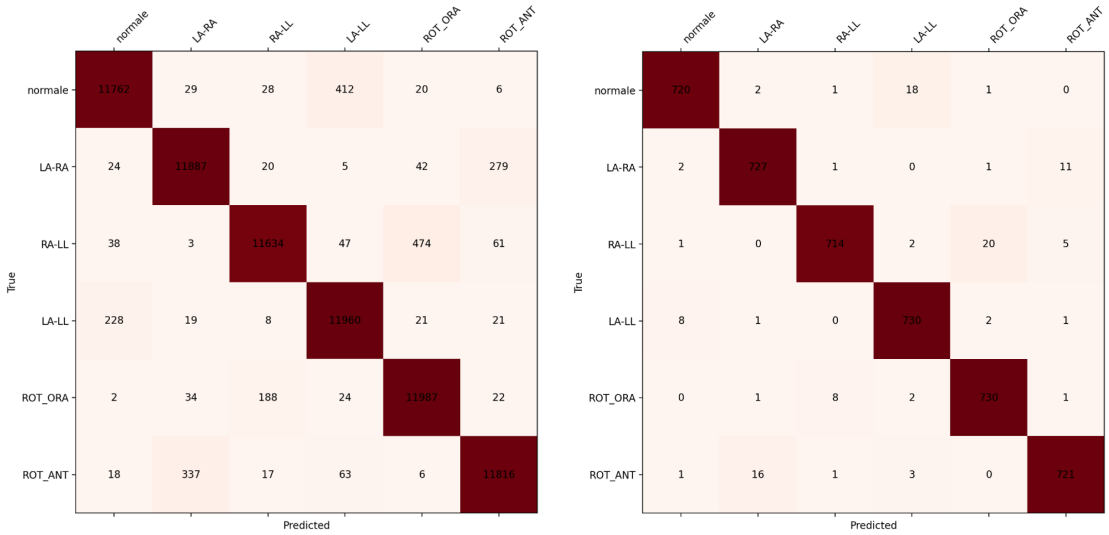
Table 4.2: Per-class performance on limb leads (simulated-to-simulated scenario, patient level).

Class	Acc.	Prec.	Rec.	Spec.	F1
normal	0.9924	0.9836	0.9704	0.9968	0.9769
LA–RA	0.9921	0.9732	0.9798	0.9946	0.9765
RA–LL	0.9912	0.9848	0.9623	0.9970	0.9734
LA–LL	0.9917	0.9669	0.9838	0.9933	0.9753
ROT_ORA	0.9919	0.9682	0.9838	0.9935	0.9759
ROT_ANT	0.9912	0.9756	0.9717	0.9951	0.9737

All classes maintain F1-scores above 0.973, with per-class accuracy always greater than 0.991. In the binary normal vs abnormal analysis, accuracy is 0.9924, with sensitivity equal to 0.9968, specificity 0.9704, precision 0.9941 and F1-score 0.9954. The binary confusion matrix in figure 4.1 shows only 12 abnormal patients erroneously classified as normal and 22 normal patients labelled abnormal.

4.1.2 Effect of Semi-Supervised Training

To evaluate the impact of semi-supervised training (SSL) on classification performance, we consider a scenario in which the model was retrained from scratch using the semi-supervised training set described in Section 3.3.3. Tables 4.3 and 4.4 report aggregate metrics at window and patient level, comparing the baseline model with the SSL-initialized model.



Window-level analysis. Patient-level analysis.

Figure 4.1: Confusion matrices for limb leads, simulated-to-simulated scenario.

Table 4.3: Aggregate performance on limbs simulated-to-simulated (window level): baseline vs SSL model.

Metric	Baseline	SSL	Δ
Accuracy	96.61%	97.22%	+0.61 pp
AUROC	0.9969	0.9973	+0.0004
AuPRC	0.9886	0.9900	+0.0014

At window level, SSL model achieved an accuracy of 97.22%, compared to the baseline 96.61%, while AUROC and AuPRC rise from 0.9969 to 0.9973 and from 0.9886 to 0.9900, respectively. Per-class metrics show F1-scores between 0.9705 and 0.9732 for all classes.

In the corresponding binary analysis, accuracy reaches 0.9902, with specificity equal to 0.9792, sensitivity 0.9924, precision 0.9958 and F1-score 0.9941.

At patient level, the improvement is smaller but still positive, as summarized in Table 4.4.

Table 4.4: Aggregate performance on limbs simulated-to-simulated (patient level): baseline vs SSL model.

Metric	Baseline	SSL	Δ
Accuracy	97.53%	97.82%	+0.29 pp
AUROC	0.9969	0.9972	+0.0003
AuPRC	0.9889	0.9899	+0.0010

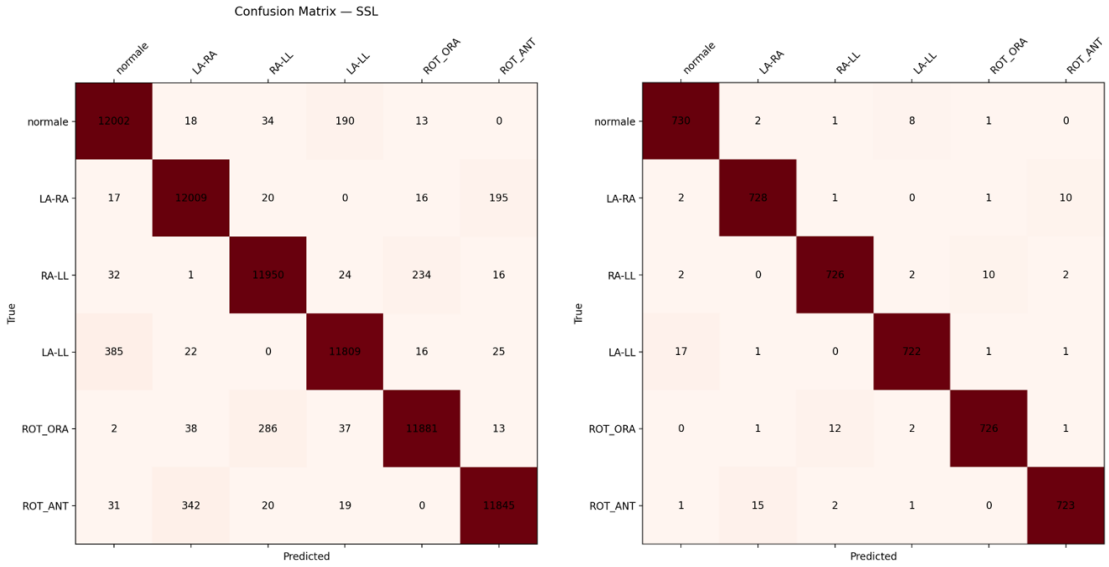
The SSL model maintains per-class F1-scores between 0.9772 and 0.9804. In binary configuration, accuracy is 0.9924, with specificity 0.9838, sensitivity 0.9941, precision 0.9968 and F1-score 0.9954; the confusion matrix in figure 4.2 shows only 22 false negatives and 12 false positives in more than 4 400 records. Consequently, the semi-supervised model achieved higher specificity and precision but lower sensitivity, while binary accuracy and F1-score remained unchanged.

4.1.3 Precordial Leads: Simulated-to-Simulated Scenario

Window level

For precordial leads, at window level, the overall accuracy is 98.04%, with AUROC equal to 0.9989 and AuPRC equal to 0.9928. Detailed per-class metrics for the 16 configurations are reported in Table 4.5. These values are slightly higher than the ones obtained in the corresponding limb lead evaluation.

Almost all classes have F1-scores above 0.97, with accuracy often close to 1. In the binary normal vs abnormal analysis, binary accuracy is 0.9906, with specificity 0.9404, sensitivity 0.9940, precision 0.9960 and F1-score 0.9950. The binary confusion matrix reports in figure 4.3 196 528 abnormal windows correctly classified and 12 395 normal windows correctly classified, with 1 187 false negatives and 786 false positives.



Window-level analysis. Patient-level analysis.

Figure 4.2: Confusion matrices for limb leads, SSL scenario.

Table 4.5: Per-class performance on precordial leads (simulated-to-simulated scenario, window level).

Class	Acc.	Prec.	Rec.	Spec.	F1
0	0.9906	0.9126	0.9404	0.9940	0.9263
1	0.9974	0.9736	0.9848	0.9982	0.9792
2	0.9992	0.9933	0.9936	0.9995	0.9934
3	0.9996	0.9961	0.9976	0.9997	0.9969
4	0.9995	0.9960	0.9964	0.9997	0.9962
5	0.9997	0.9988	0.9969	0.9999	0.9978
6	0.9944	0.9644	0.9447	0.9977	0.9544
7	0.9989	0.9943	0.9887	0.9996	0.9915
8	0.9985	0.9827	0.9939	0.9988	0.9883
9	0.9993	0.9966	0.9925	0.9998	0.9946
10	0.9978	0.9838	0.9804	0.9989	0.9821
11	0.9980	0.9783	0.9904	0.9985	0.9844
12	0.9991	0.9963	0.9898	0.9998	0.9931
13	0.9967	0.9734	0.9744	0.9982	0.9739
14	0.9974	0.9827	0.9756	0.9989	0.9791
15	0.9945	0.9656	0.9464	0.9978	0.9559

Patient level

At record level, the overall accuracy reaches 98.49%, with AUROC equal to 0.9992 and AuPRC equal to 0.9948. The per-class metrics reported in Table 4.6 show F1-scores generally above 0.98, with some classes achieving F1 equal to 1.0000.

Table 4.6: Per-class performance on precordial leads (simulated-to-simulated scenario, patient level).

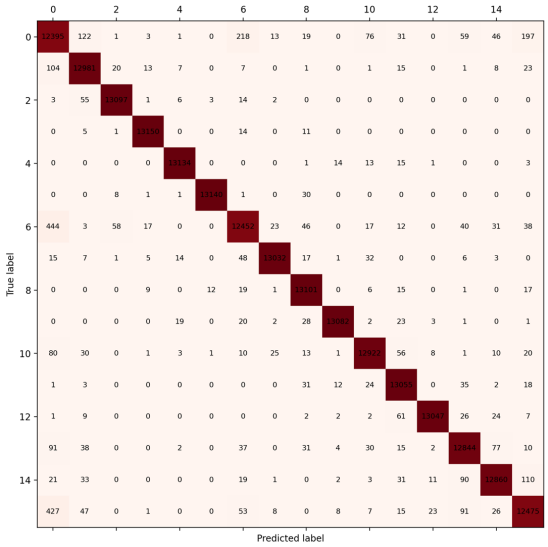
Class	Acc.	Prec.	Rec.	Spec.	F1
0	0.9927	0.9328	0.9517	0.9954	0.9422
1	0.9982	0.9821	0.9888	0.9988	0.9854
2	0.9994	0.9944	0.9955	0.9996	0.9949
3	1.0000	1.0000	1.0000	1.0000	1.0000
4	0.9998	0.9989	0.9978	0.9999	0.9983
5	0.9998	0.9989	0.9978	0.9999	0.9983
6	0.9960	0.9760	0.9596	0.9984	0.9677
7	0.9993	0.9944	0.9944	0.9996	0.9944
8	0.9992	0.9911	0.9955	0.9994	0.9933
9	0.9995	0.9977	0.9944	0.9999	0.9961
10	0.9981	0.9865	0.9831	0.9991	0.9848
11	0.9982	0.9800	0.9910	0.9987	0.9855
12	0.9994	0.9989	0.9921	0.9999	0.9955
13	0.9971	0.9732	0.9809	0.9982	0.9771
14	0.9978	0.9864	0.9775	0.9991	0.9819
15	0.9954	0.9682	0.9584	0.9979	0.9633

In the corresponding binary configuration, accuracy is 0.9927, with sensitivity 0.9954, specificity 0.9517, precision 0.9968 and F1-score 0.9961. The confusion matrix in figure 4.3 shows 13 289 abnormal patients correctly classified, 61 false negatives, 847 true normal patients and 43 false positives.

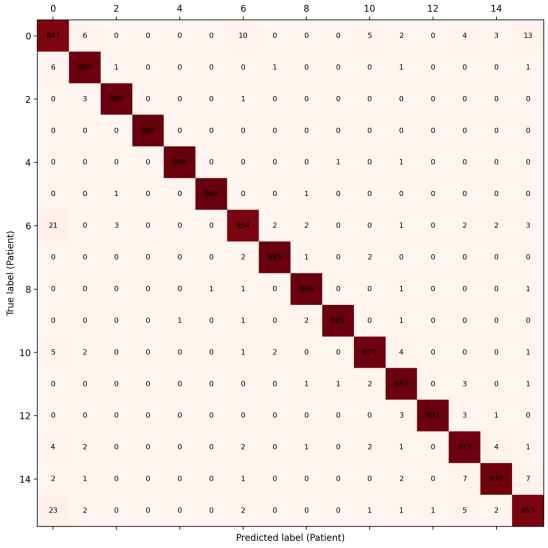
4.2 Simulated-to-Real Scenario: 6-lead and 7-lead Models

The final scenario considers the transfer of models trained on simulated data to a real clinical dataset, restricted to limb exchanges. Two approaches are compared: a baseline 6-lead model, based only on limb leads, and a 7-lead model obtained by using the same six leads and adding the precordial electrode **V6**.

In order to perform a fair comparison, both models were trained from scratch



Window-level analysis.



Patient-level analysis.

Figure 4.3: Confusion matrices for precordial leads.

using the same training dataset and the same optimization strategy. The purpose of this experiment is to assess whether the additional information provided by V6 improves generalization to real clinical ECGs.

Aggregate performance is reported in Table 4.7. Adding the seventh lead increases **accuracy** from 88.58% to 90.44%, and improves AuPRC from 0.7384 to 0.7564, while AUROC decreases slightly from 0.9596 to 0.9465.

Table 4.7: Aggregate performance in the simulated-to-real scenario at ECG level for the 6-lead and 7-lead configurations.

Model	Accuracy	AUROC	AuPRC
6 leads	88.58%	0.9596	0.7384
7 leads	90.44%	0.9465	0.7564

Compared with the 6-lead configuration, the inclusion of V6 increases accuracy by 1.86 percentage points and AuPRC by 0.0180, while AUROC decreases by 0.0131.

The total number of misclassified ECG cases decreases from 49 for the 6-lead configuration to 41 for the 7-lead configuration, corresponding to a net reduction of eight errors.

To make this behaviour more explicit, Table 4.8 reports, for each true class, the number of patients and the number of errors with and without V6. The improvement is clearly non-homogeneous: some classes benefit substantially, others show marginal impact, and a few even worsen.

Table 4.8: Class-wise errors in the simulated-to-real evaluation. Change is computed as the number of 6-lead errors minus the number of 7-lead errors; positive values indicate fewer errors with the 7-lead configuration.

True class	ECG cases	Errors 6-lead	Errors 7-lead	Change
Normal	191	7	3	+4
LA–LL	13	6	10	-4
LA–RA	178	6	4	+2
ROT_ANT	26	21	20	+1
RA–LL	14	3	3	0
ROT_ORA	7	6	1	+5
Total	429	49	41	+8

Per-class patient-level metrics are reported in Table 4.9. The *normal* and *LA–RA* classes maintain F1-scores above 91% in both models. In contrast, classes such

as *LA–LL* and *ROT_ANT* remain significantly more problematic.

Table 4.9: Per-class metrics in the simulated-to-real scenario (patient level).

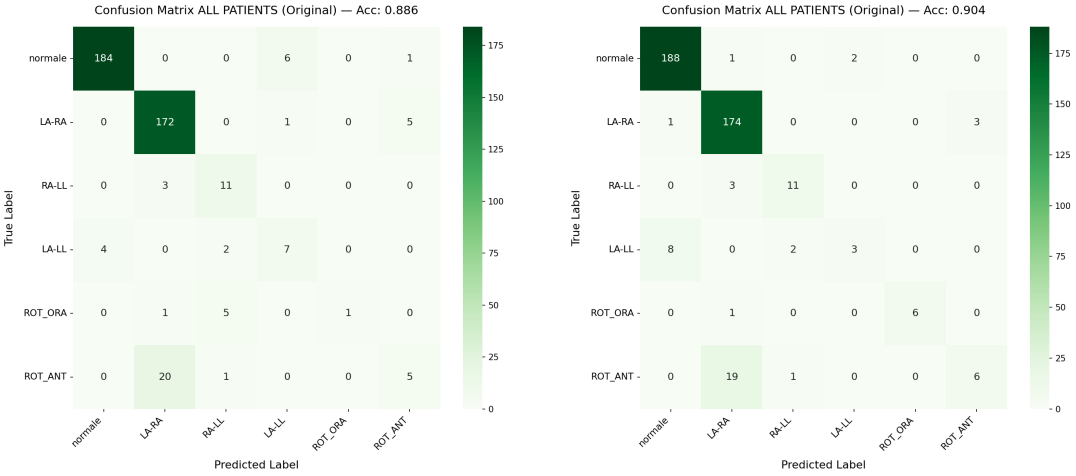
6-lead model					
Class	Acc.	Prec.	Rec.	Spec.	F1
normal	97.44%	97.87%	96.34%	98.32%	97.10%
LA–RA	93.01%	87.76%	96.63%	90.44%	91.98%
RA–LL	97.44%	57.89%	78.57%	98.07%	66.67%
LA–LL	96.97%	50.00%	53.85%	98.32%	51.85%
ROT_ORA	98.60%	100.00%	14.29%	100.00%	25.00%
ROT_ANT	93.71%	45.45%	19.23%	98.51%	27.03%
7-lead model					
Class	Acc.	Prec.	Rec.	Spec.	F1
normal	97.20%	95.43%	98.43%	96.22%	96.91%
LA–RA	93.47%	87.88%	97.75%	90.44%	92.55%
RA–LL	98.60%	78.57%	78.57%	99.28%	78.57%
LA–LL	97.20%	60.00%	23.08%	99.52%	33.33%
ROT_ORA	99.77%	100.00%	85.71%	100.00%	92.31%
ROT_ANT	94.64%	66.67%	23.08%	99.26%	34.29%

The most evident effect of adding **V6** is observed in the **ROT_ORA** class, whose F1-score increases from 25.00% to 92.31%. Improvements for *RA–LL* are also noticeable, whereas *LA–LL* and *ROT_ANT* still show marked difficulties.

Despite the complexity of the multi-class problem, binary *normal vs abnormal* performance at patient level remains very high in both models. For the 6-lead model, binary accuracy is 97.90%, with sensitivity 98.74%, specificity 96.86%, precision 97.51% and F1-score 98.12%. For the 7-lead model, binary accuracy is 97.20%, with sensitivity 96.22%, specificity 98.43%, precision 98.71% and F1-score 97.45%. The corresponding confusion matrices in Figure 4.4 show very few normal patients misclassified as abnormal and vice versa in both cases.

4.2.1 Lightweight Complexity Reduction

Once the role of **V6** in salvable patients and the impact of manual labelling are clarified, the next step is to evaluate how much the input configuration can be simplified without losing essential information. The idea is to check whether, after ensuring at least one effective precordial lead, it is still necessary to use all peripheral leads, or whether more compact but still functional setups exist.



Pathology model with six limb leads. Pathology model with the addition of V6.

Figure 4.4: Confusion matrices in the simulated-to-real scenario for the 6-lead and 7-lead models, trained on the pathology dataset.

Three main configurations are compared:

- **limbs_extended** (III, aVR, V6): two independent peripheral leads on the frontal plane plus V6 as three-dimensional component.
- **limbs_complete** (all peripheral leads + V6): reference setup with maximum accuracy.
- **limbs_v1+v6** (all peripheral leads + V6 + V1): used to assess whether adding an extra precordial lead provides real benefit or not.

Table 4.10: Patient-level accuracy for different lead configurations.

Model / setup	Input leads	Accuracy	Note
limbs_extended	III, aVR, V6	0.897	lightweight setup, frontal basis + V6
limbs_complete	all peripheral leads + V6	0.904	most accurate configuration
limbs_v1+v6	complete + V1	0.883	extra complexity without clear benefit

The **limbs_complete** configuration achieves the best performance, but the **limbs_extended** setup remains surprisingly close, with only a small loss of accuracy. This suggests that two appropriately chosen peripheral leads are sufficient to define the **vector basis** of the frontal plane, while **V6** adds the missing spatial component needed to resolve many ambiguities. In contrast, adding V1 in **limbs_v1+v6** does not bring proportional improvement, and the accuracy drop points to possible information saturation and higher sensitivity to noise.

In summary, the **2 limb + V6** combination emerges as a **lightweight but functional** model: it preserves most of the ability to recover salvable patients, reduces acquisition system complexity, and maintains accuracy very close to the complete setup. From an application perspective, this solution offers a promising compromise between simplicity, robustness and geometric coherence, especially considering that part of the residual error is more related to manual labelling limits than to the model's ability to interpret the ECG signal.

4.3 Summary of Results and Findings

On simulated data, both limb lead and precordial lead models achieved high performance, with AUROC values above 0.996 and binary F1-scores of approximately 0.99. The semi-supervised approach slightly increased the performance for the multi-class task at both window level and ECG level.

In the sim-to-real scenario, multi-class accuracy was lower with respect to the simulated evaluation. Binary normal-versus-abnormal performance scored F1-scores above 97% for both the 6-lead and 7-lead configurations. The 7-lead configuration achieved an higher multi-class accuracy and AuPRC with respect to the 6-lead configuration, while binary sensitivity and binary F1-score remained lower.

The interpretation of the simulated-to-real performance, the role of domain shift, the contribution of V6, and the clinical implications of the observed sensitivity–specificity trade-offs are discussed in Chapter 5.

Chapter 5

General Discussion

This chapter discusses the experimental results presented in the previous Chapter 4, relating them to the objectives of the thesis, the methodological choices introduced earlier, and the practical requirements of automatic ECG electrode-interchange detection. The results should be interpreted on two complementary levels. On the one hand, the simulated-to-simulated experiments assess whether the proposed models are able to learn the transformations associated with the considered interchange classes under controlled conditions. Meanwhile, the simulated-to-real evaluation examines whether the information learned from simulated data remains useful when the models are applied to clinical ECGs, where biological variability, pathological patterns, acquisition artifacts, and uncertainty in the reference labels all play a role.

In general, three main findings emerge from the experiments. First, both the limb-lead and precordial-lead models achieve high performance on fully simulated data, and aggregating predictions at the ECG level improves the results compared to individual windows. Second, the semi-supervised extension leads to a modest improvement in multi-class performance, but also alters the balance between sensitivity and specificity in the binary task. Third, when moving to real clinical ECGs, there is a clear drop in fine-grained multi-class performance, while the binary distinction between normal and abnormal configurations remains relatively strong. The contribution of V6 is not uniform: it provides clear benefits for some configurations, but it does not resolve all the difficulties associated with every interchange class.

5.1 Performance in the Simulated-to-Simulated Setting

The high performance obtained on simulated limb-lead and precordial-lead data indicates that the proposed architectures are capable of learning the transformations used to generate the different electrode-interchange classes. In this controlled setting, both training and test samples are generated using the same mathematical or lead-exchange rules, and preprocessing steps such as sampling frequency, window length, and normalization are consistent across datasets. As a result, the task is well-defined and does not include the full range of uncertainties typically found in clinical recordings.

For limb leads, the relatively narrow spread of class-wise F1-scores suggests that the LDenseNet does not concentrate its performance only on the most frequent or easiest classes. This is relevant because the task includes both pairwise interchanges and rotational configurations, which produce different transformations in the frontal plane. The high sensitivity observed in the binary setting further shows that abnormal configurations are rarely classified as normal in the simulated domain. Since the intended application prioritizes the detection of any possible interchange before identifying its exact type, this result supports the validity of the binary formulation used in the evaluation protocol.

The precordial model achieves slightly higher aggregate performance than the limb-lead model. The ILC architecture explicitly combines learned single-lead representations with inter-lead correlation coefficients, and this design aligns well with the sequential spatial organization of V1–V6. This observation is also consistent with the motivation of the original ILC model, where shared branch parameters maintain a common feature space across leads and correlation terms capture deviations in expected inter-lead relationships Huang et al. [2024]. However, strong aggregate performance should not be interpreted as evidence that all precordial swaps are equally easy. Adjacent precordial leads often exhibit similar morphologies, making interchanges between neighbouring positions more difficult than those involving distant leads. This aspect should be considered when analysing class-wise results, and it supports the choice of reporting each precordial interchange explicitly rather than relying only on numerical labels.

Overall, the simulated results mainly demonstrate that the models and data-generation procedures are internally consistent. They confirm that the architectures can recover the simulated class structure and provide a necessary validation step before testing on clinical data. However, they do not, by themselves, demonstrate generalization, since the same simulation rules are used for both training and testing.

5.2 Effect of ECG-level Probability Aggregation

For both limb and precordial leads, performance at the ECG level is higher than at the window level. The final ECG prediction is obtained by averaging the probability vectors produced for all valid windows and selecting the class with the highest mean probability. This soft-voting approach reduces the influence of individual windows affected by transient artifacts, local morphological ambiguity, or incomplete cardiac cycles.

The improvement observed after aggregation suggests that many window-level errors are not consistent throughout the entire ECG. A window containing noise or atypical beats may contribute to the prediction, but does not dominate the final decision if the remaining windows provide stable evidence for another class. This behaviour is appropriate for electrode interchange, which is a recording-level condition and should generally affect the entire signal rather than isolated segments.

At the same time, ECG-level samples derived from simulated data should not be interpreted as independent clinical cases. When multiple interchange configurations are generated from the same original ECG, they share the same underlying cardiac morphology. As a consequence, the number of evaluated samples exceeds the number of independent subjects. While the reported ECG-level metrics remain valid for the chosen evaluation unit, this dependency should be considered when comparing results with studies that report one observation per patient.

5.3 Contribution of the Semi-Supervised Extension

The semi-supervised extension was introduced to leverage clinical ECGs without reliable manual labels. High-confidence pseudo-labels generated by the baseline model are combined with simulated training data, and the signal-quality procedure is applied again to reduce the inclusion of invalid recordings. This strategy leads to a modest improvement in aggregate multi-class accuracy, AUROC, and AuPRC at both window and ECG levels.

One possible explanation is that pseudo-labelled clinical ECGs introduce types of variability that are not well represented in purely simulated data. Real recordings include pathological morphologies, patient-specific electrical axes, acquisition differences, and residual noise after preprocessing. Even if their labels are automatically generated, exposure to this variability may help the model learn features that are less dependent on idealized simulated patterns.

The magnitude of the improvement remains limited, which is expected. The baseline model already performs close to saturation on simulated data, leaving limited room for further gains. Moreover, pseudo-labelled data do not provide

independent ground truth: they are selected based on the confidence of a model trained on simulated labels. This may reinforce existing decision boundaries and propagate systematic errors, especially in geometrically ambiguous cases. A high confidence threshold reduces this risk, but cannot eliminate it, since confidence does not guarantee correctness.

The binary results also show that the effect of the semi-supervised extension is not uniformly beneficial. Specificity and precision increase, while sensitivity decreases. At the ECG level, the total number of errors and the F1-score remain unchanged, but errors shift from false positives to false negatives. This trade-off is important in practice. Reducing false alarms may improve usability, especially because correct electrode placement is much more common than interchange. However, failing to detect an abnormal ECG may be more critical, as the technical error could remain unnoticed and affect clinical interpretation. For this reason, model selection should not rely only on overall accuracy, but should also consider the relative cost of different types of errors.

5.4 Generalization from Simulated to Real Clinical ECGs

The simulated-to-real evaluation shows a substantial decrease in multi-class performance compared to the fully simulated experiments. This result is consistent with a domain shift between mathematically generated interchanges and real clinical recordings. Simulated transformations preserve signal quality and apply exact known operations, whereas real ECGs combine electrode misplacement with pathology, variability in cardiac axis, anatomical differences, baseline instability, muscular noise, acquisition artifacts, and uncertainty in clinical annotations.

The class-wise results indicate that this domain shift does not affect all classes equally. Normal and LA–RA configurations remain relatively well recognized, whereas LA–LL and ROT_ANT are considerably more challenging. Part of this difference may be due to class imbalance: the real test set contains many more normal and LA–RA cases than examples of other configurations. Metrics computed on very small sample sizes are inherently unstable, as a small number of misclassifications can lead to large changes in performance measures.

Despite the drop in multi-class performance, the model does not fail completely on real data. Binary normal-versus-abnormal F1-scores remain above 97%, indicating that the presence of an abnormal configuration is often detected even when the exact class is not correctly identified. This distinction is clinically relevant. A system used for quality control or screening may still be useful if it reliably triggers a review or repeated acquisition, even when the precise type of interchange is uncertain.

However, these results are not sufficient to claim clinical readiness. The real dataset is limited, some classes are under-represented, and the reliability of reference labels is uncertain. The system should therefore be considered a decision-support or quality-control tool rather than an autonomous diagnostic system.

5.5 Geometric Interpretation of the Contribution of V6

The rationale for including V6 is related to the geometric limitations of limb leads, which represent projections of cardiac electrical activity on the frontal plane and are therefore linearly redundant. Different combinations of cardiac axis and electrode interchange can produce similar or even indistinguishable patterns in these leads, as discussed in terms of vector mimicry and collapse conditions.

V6 introduces an additional spatial projection that is not redundant with the limb-lead system. While it does not reconstruct full three-dimensional cardiac activity, it can preserve morphological differences that are not visible in the frontal plane. In some cases, configurations that appear similar in limb leads can still be distinguished when V6 is included. This interpretation is consistent with the observed improvements for certain classes and with the reduction in overall multi-class errors when using seven leads.

At the same time, the benefit is not uniform. Some classes do not improve, and certain configurations remain difficult. This suggests that V6 can help to resolve some ambiguities, but it is not a universal solution. In cases close to theoretical collapse conditions, different configurations may remain indistinguishable across all available leads. In such situations, the limitation is due to the available information rather than the model.

It is important to note that the comparison between the two configurations is fair, since it was performed using the same training dataset and the same optimization strategy. This allows to associate the performance changes to the information provided by V6.

5.6 Error Interpretation and Uncertainty of the Clinical Ground Truth

The analysis of errors in real data suggests at least three different sources. The first includes cases that are potentially recoverable, where ambiguity in limb leads could be resolved by additional information. The second includes structurally ambiguous cases, where different configurations produce very similar signals. The

third involves recordings affected by poor signal quality, pathological patterns, or uncertainty in the assigned labels.

The signal-quality assessment included in the preprocessing pipeline represents an important strength, as it removes clearly invalid or corrupted segments before classification. However, no heuristic procedure can eliminate all artifacts, and borderline cases may still pass through the filters.

Uncertainty in ground truth is particularly relevant for some classes. Certain pathological conditions can mimic electrode interchanges, and conversely, some interchanges may go unnoticed during routine clinical reporting. While qualitative inspection can highlight doubtful cases, it cannot be used to review labels without introducing bias.

For this reason, performance metrics should remain based on the original annotations. Any alternative evaluation would require an independent expert adjudication process, ideally involving multiple clinicians and blinded review. Without this, label uncertainty should be acknowledged as a limitation rather than a corrected post hoc.

5.7 Clinical Interpretation of the Binary Task

The binary formulation corresponds to a practical clinical question: whether an ECG should be accepted or reviewed. In this context, sensitivity reflects the ability to detect abnormal recordings, while specificity reflects the ability to avoid false alarms.

The observed sensitivity–specificity trade-off highlights the need to adapt the operating point to the intended application. A fully automatic rejection system should prioritize sensitivity, whereas a decision-support tool may tolerate more false positives.

The strong binary performance observed on real data suggests potential for use as a screening or quality-control tool. However, prospective validation is required to assess its impact in real workflows.

5.8 Lightweight Input Configurations

The reduced configuration using III, aVR, and V6 introduced in section 4.2.1 achieves performance close to the one of the full configuration. This is consistent with the redundancy of limb leads, where a limited subset can capture most of the relevant information.

From a practical perspective, reducing the number of input leads decreases computational cost and facilitates deployment in resource-constrained environments.

However, the effectiveness of this configuration depends on the correct placement of V6. If V6 itself is misplaced, the model may lose its main source of additional information or be deceived. This suggests that future systems should include mechanisms to assess the reliability of precordial leads.

5.9 Methodological Strengths and Limitations

This work presents several methodological strengths. These include a structured preprocessing pipeline with explicit signal-quality assessment, evaluation at both window and ECG levels, and the combination of simulated and real data. In addition, the analysis explicitly considers the geometric limitations of the ECG leads rather than attributing all errors to the model.

Despite this, several limitations should be acknowledged. Reliance on simulated data limits the realism of the training process. The clinical dataset is relatively small and unbalanced, which affects the reliability of class-wise metrics. The semi-supervised approach depends on pseudo-labels that may introduce bias. Furthermore, the precordial model has not yet been validated on a sufficiently large real dataset, and uncertainty in clinical annotations remains an important issue.

5.10 Overall Interpretation

The results show that lightweight deep learning models can effectively learn electrode-interchange patterns under controlled conditions. However, generalization to real clinical data remains challenging, particularly for detailed multi-class classification.

The inclusion of V6 provides useful additional information in some cases, but it does not fully resolve all ambiguities. The binary detection task appears to be more robust and currently represents the most promising application.

Overall, the proposed system is better suited as a quality-control and decision-support tool rather than as a fully autonomous diagnostic solution. Its practical value will depend on maintaining high sensitivity, managing false alarms, integrating signal-quality assessment, and validating the complete pipeline in real clinical settings.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

The aim of this work was to develop a unified system that automatically detects ECG electrode interchange using lightweight deep learning models. This was done through a complete pipeline, composed of: signal preprocessing, signal quality assessment, simulation of electrode interchange, window-level classification and ECG-level prediction. LDenseNet was the model developed for the limb electrode interchange task, while for the precordial interchange task, the Inter Lead Correlation model was implemented. A further evaluation was done by applying a semi supervised approach, based on high confidence pseudo labelled clinical ECGs.

Results on simulated data show that the models achieved high performance both in the multi-class and binary scenarios. Further improvement was recorded aggregating the results at ECG level, while the semi supervised extension produced a modest improvement in the multi-class results.

The simulated to real scenario highlighted one of the main difficulties of this work: the model correctly distinguishes between normal and abnormal ECGs, but with difficulty at recognising the specific interchange configuration in clinical ECGs, also due to misleading classes. Including V6, the model achieved higher multi-class accuracy and AuPRC than the six limb lead configuration, although it did not resolve all geometric ambiguities. Then, results of the reduced configuration (leads III, aVR and V6) showed performance close to that of the complete configuration, while reducing complexity.

Overall, the proposed pipeline provides a promising basis for an automatic ECG quality-control system that could support specialists in identifying recordings affected by electrode interchange errors. Moreover, the limited complexity of the

lightweight models used, makes them suitable for portable and wearable ECG systems. This is particularly relevant for the portable acquisition system developed by Aizoi, although this deployment would require further evaluation on computational requirements and real time performance.

6.2 Future Work

A future work could include the study of precordial misplacement, that in this thesis was only mentioned for completeness. This task could be performed by building a dedicated clinical dataset, which can be subsequently used to evaluate the ILC model under more realistic conditions.

A further development for the precordial misplacement task could be to reformulate the network, so it is able to predict the anatomical position of each precordial electrode, rather than assigning the ECG to a predefined interchange class. For each electrode from V1 to V6 this network could produce a separate output, indicating if the electrode is correctly placed in its standard anatomical position and, when possible, be able to detect a misplacement.

The resulting model could be integrated into the portable ECG system developed by Aizoi. In this way, evaluation on real clinical ECGs would make it possible to verify whether the architecture satisfies the computational constraints of portable acquisition and whether its predictions remain reliable under real world conditions.

Bibliography

AIZOI. Biotechware, la startup che porta l'elettrocardiogramma nel cloud, 2026. URL https://www.aizoi.com/biotechware-la-startup-che-porta-lelettrocardiogramma-nel-cloud/. Articolo su innovazione e telemedicina.

Douglas G. Altman. *Practical Statistics for Medical Research*. Chapman and Hall, London, 1994.

A. N. Author. Diagnostic testing accuracy: Sensitivity, specificity, predictive values, and likelihood ratios. In *StatPearls*. StatPearls Publishing, Treasure Island (FL), 2023.

S. Serge Barold. Einthoven's triangle transparency. *Reviews in Cardiovascular Medicine*, 11(2):129–131, 2010. doi: 10.3909/ricm0506.

Velislav N. Batchvarov, Marek Malik, and A. John Camm. Incorrect electrode cable connection during electrocardiographic recording. *Europace*, 9(11):1081–1090, 2007. doi: 10.1093/europace/eum198.

Velislav N. Batchvarov et al. Electrocardiographic lead reversals. *Indian Pacing and Electrophysiology Journal*, 23(5):xxx–xxx, 2023. doi: 10.1016/j.ipej.2023.08.004. URL https://www.sciencedirect.com/science/article/pii/S0972629223000992.

Ed Burns and Robert Buttner. ECG limb lead reversal, 10 2024. URL https://litfl.com/ecg-limb-lead-reversal-ecg-library/. Accessed: 2024-03-05.

Tim De Coster, Arman Nobacht, Thom Oostendorp, Antoine A. F. de Vries, Ruben Coronel, and Daniël A. Pijnappels. Monitoring and modulating cardiac bioelectricity: from einthoven to end-user. *Europace*, 27(1):euae300, 2025. doi: 10.1093/europace/euae300.

Dreamstime. Chest ecg leads placement illustration - six colored electrocardiography cables, 2026. URL https://it.dreamstime.com/illustrazione-di-stock-il-ecg-del-petto-conduce-l-illustrazione-di-disposizio Immagine stock vettoriale.

Javier García-Niebla, Patricia Llontop-García, José Ignacio Valle-Racero, Guillem Serra-Autonell, Antoni Bayés de Luna, and Miguel Fiol-Sala. Interpreting electrocardiograms in the presence of limb lead reversal: The importance of the p wave. *Journal of Electrocardiology*, 47(6):795–798, 2014. doi: 10.1016/j.jelectrocard.2014.06.002.

Karim Hajian-Tilaki. Receiver-operating characteristic (roc) curve analysis for medical diagnostic test evaluation. *Biochemia Medica*, 23(2):157–162, 2013.

E. William Hancock, Marc Roelke, Rolf Childers, et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram. Part V: Conduction disorders. *Journal of the American College of Cardiology*, 53(11):992–1002, 2009. doi: 10.1016/j.jacc.2008.12.014.

Yangcheng Huang, Mingjie Wang, Yi-Gang Li, and Wenjie Cai. A lightweight deep learning approach for detecting electrocardiographic lead misplacement. *Physiological Measurement*, 45(5):055006, 2024. doi: 10.1088/1361-6579/ad43ae.

Klaus Ilg and Georg Lehmann. The effect of precordial lead displacement on ecg morphology. *Journal of Electrocardiology*, 45(6):672–675, 2012. doi: 10.1016/j.jelectrocard.2012.02.007.

Irena Jekova, Vessela Krasteva, Remo Leber, et al. Inter-lead correlation analysis for automated detection of cable reversals in 12/16-lead ecg. *Computer Methods and Programs in Biomedicine*, 134:31–41, 2016. doi: 10.1016/j.cmpb.2016.06.007.

Mikolaj Kania, Rohini Senthilnathan, Anish N. Bhuva, et al. Magnetocardiography for identification of coronary ischemia in patients with chest pain and normal resting 12-lead electrocardiogram. *Annals of Noninvasive Electrocardiology*, 25 (4):e12715, 2020. doi: 10.1111/anec.12715.

Mohammad Kayyali, Ana Mincholé, Shuang Qian, Alistair Young, Devran Ugurlu, Elliot Fairweather, et al. Anatomical-electrical coupling of cardiac axes: Definitions and population variability for advancing personalised ecg interpretation. *PLoS Computational Biology*, 21(7):e1013161, 2025. doi: 10.1371/journal.pcbi.1013161.

Donghyeon Kim, Kyoung Ryul Lee, Dong Seok Lim, Kwang Hyun Lee, Jong Seon Lee, Dae-Yeol Kim, and Chae-Bong Sohn. A novel hybrid cnn-transformer model for arrhythmia detection without r-peak identification using stockwell transform. *Scientific Reports*, 15(1):7817, 2025. doi: 10.1038/s41598-025-92582-9.

Paul Kligfield, Leonard S. Gettes, James J. Bailey, et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram. Part I: The electrocardiogram and its technology. *Circulation*, 115(10):1306–1324, 2007. doi: 10.1161/CIRCULATIONAHA.106.180200.

Lisa Kuhn and Louise Rose. Ecg interpretation part 1: Understanding mean electrical axis. *Journal of Emergency Nursing*, 34(5):415–418, 2008. doi: 10.1016/j.jen.2008.06.002.

Peter W. Macfarlane, Adriaan van Oosterom, Olle Pahlm, Paul Kligfield, and Michiel J. Janse. *Comprehensive Electrocardiology*. Springer London, London, 2nd edition, 2010. ISBN 9781848820455. doi: 10.1007/978-1-84882-045-5.

Margaret S. Pepe. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, Oxford, 2003.

Esther Pueyo, Pablo Laguna, and Juan Pablo Martínez. Techniques for ventricular repolarization instability assessment from the ECG. *Proceedings of the IEEE*, 104(2):392–415, 2016. doi: 10.1109/JPROC.2015.2500501.

Alireza Rafiei, Trisha Dwivedi, Joel Xue, et al. Toward self-validating ecg systems: A personalized, uncertainty-aware approach for detecting and estimating lead misplacement. *Journal of Electrocardiology*, 95:154204, 2026. doi: 10.1016/j.jelectrocard.2025.01.001.

Stephen J. Redmond, Nigel H. Lovell, Jim Basilakis, and Branko G. Celler. Ecg quality measures in telecare monitoring. In *30th Annual International IEEE EMBS Conference*, pages 2869–2872, 2008. doi: 10.1109/IEMBS.2008.4650054.

Khaled Rjoob, Raymond Bond, Dewar Finlay, et al. Machine learning techniques for detecting electrode misplacement and interchanges when recording ecgs: A systematic review and meta-analysis. *Journal of Electrocardiology*, 62:116–123, 2020. doi: 10.1016/j.jelectrocard.2020.07.014.

Khaled Rjoob, Raymond Bond, Dewar Finlay, Victoria McGilligan, et al. Reliable deep learning-based detection of misplaced chest electrodes during electrocardiogram recording: Algorithm development and validation. *JMIR Medical Informatics*, 9(4):e25347, 2021. doi: 10.2196/25347.

Udit Satija, Barathram Ramkumar, and M. Sabarimalai Manikandan. A review of signal processing techniques for electrocardiogram signal quality assessment. *IEEE Reviews in Biomedical Engineering*, 11:36–52, 2018. doi: 10.1109/RBME.2018.2810159.

Ana-Maria Šimundić. Simple statistical measures for diagnostic accuracy assessment. *Journal of Clinical Pathology*, 63(12):1037–1039, 2010.

Borys Surawicz and Timothy K. Knilans. *Chou's Electrocardiography in Clinical Practice*. Elsevier, Philadelphia, 6th edition, 2008. ISBN 9781416036163. doi: 10.1016/B978-1-4160-3616-3.X5001-3.

Borys Surawicz, Rolf Childers, Barbara J. Deal, Leonard S. Gettes, et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram. Part III: Intraventricular conduction disturbances. *Journal of the American College of Cardiology*, 53(11):976–981, 2009. doi: 10.1016/j.jacc.2008.12.013.

A. S. Vink, S. A. B. Clur, N. A. Blom, et al. Effect of age and gender on the QTc interval in healthy individuals. *The American Journal of Cardiology*, 122 (8):1302–1308, 2018. doi: 10.1016/j.amjcard.2018.06.031.

Wikimedia Commons. File:limb leads of ekg.png — wikimedia commons, the free media repository, 2026. URL https://commons.wikimedia.org/wiki/File:Limb_leads_of_EKG.png. Diagramma delle derivazioni periferiche e del sistema esassiale.

Wikipedia contributors. Willem einthoven — wikipedia, l'enciclopedia libera, 2026. URL https://it.wikipedia.org/wiki/Willem_Einthoven. Versione italiana.

Jialing Xie, Li Peng, Liang Wei, Yushun Gong, et al. A signal quality assessment-based ecg waveform delineation method used for wearable monitoring systems. *Medical & Biological Engineering & Computing*, 59:2073–2084, 2021. doi: 10.1007/s11517-021-02425-8.