

## Article

# Delineation of 12-Lead ECG Representative Beats Using Convolutional Encoder–Decoders with Residual and Recurrent Connections

Vessela Krasteva <sup>1</sup>, Todor Stoyanov <sup>1</sup>, Ramun Schmid <sup>2</sup> and Irena Jekova <sup>1,\*</sup>

<sup>1</sup> Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Acad. G. Bonchev Str. Bl. 105, 1113 Sofia, Bulgaria; vessika@biomed.bas.bg (V.K.); todor@biomed.bas.bg (T.S.)

<sup>2</sup> Signal Processing, Schiller AG, Altgasse 68, CH-6341 Baar, Switzerland; ramun.schmid@schiller.ch

\* Correspondence: irena@biomed.bas.bg

**Abstract:** The aim of this study is to address the challenge of 12-lead ECG delineation by different encoder–decoder architectures of deep neural networks (DNNs). This study compares four concepts for encoder–decoders based on a fully convolutional architecture (CED-Net) and its modifications with a recurrent layer (CED-LSTM-Net), residual connections between symmetrical encoder and decoder feature maps (CED-U-Net), and sequential residual blocks (CED-Res-Net). All DNNs transform 12-lead representative beats to three diagnostic ECG intervals (P-wave, QRS-complex, QT-interval) used for the global delineation of the representative beat (P-onset, P-offset, QRS-onset, QRS-offset, T-offset). All DNNs were trained and optimized using the large PhysioNet ECG database (PTB-XL) under identical conditions, applying an advanced approach for machine-based supervised learning with a reference algorithm for ECG delineation (ETM, Schiller AG, Baar, Switzerland). The test results indicate that all DNN architectures are equally capable of reproducing the reference delineation algorithm’s measurements in the diagnostic PTB database with an average P-wave detection accuracy (96.6%) and time and duration errors: mean values (−2.6 to 2.4 ms) and standard deviations (2.9 to 11.4 ms). The validation according to the standard-based evaluation practices of diagnostic electrocardiographs with the CSE database outlines a CED-Net model, which measures P-duration ( $2.6 \pm 11.0$  ms), PQ-interval ( $0.9 \pm 5.8$  ms), QRS-duration ( $-2.4 \pm 5.4$  ms), and QT-interval ( $-0.7 \pm 10.3$  ms), which meet all standard tolerances. Noise tests with high-frequency, low-frequency, and power-line frequency noise (50/60 Hz) confirm that CED-Net, CED-Res-Net, and CED-LSTM-Net are robust to all types of noise, mostly presenting a mean duration error < 2.5 ms when compared to measurements without noise. Reduced noise immunity is observed for the U-net architecture. Comparative analysis with other published studies scores this research within the lower range of time errors, highlighting its competitive performance.



**Citation:** Krasteva, V.; Stoyanov, T.; Schmid, R.; Jekova, I. Delineation of 12-Lead ECG Representative Beats Using Convolutional Encoder–Decoders with Residual and Recurrent Connections. *Sensors* **2024**, *24*, 4645. <https://doi.org/10.3390/s24144645>

Academic Editors: Yvonne Tran, Jan Cornelis and Carlos Sánchez

Received: 19 June 2024

Revised: 11 July 2024

Accepted: 16 July 2024

Published: 17 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** deep learning; deep neural networks; ECG signal processing; ECG segmentation; average beat; ECG interval measurements; diagnostic ECG waves

## 1. Introduction

The electrocardiogram (ECG) is a standard clinical tool for the investigation of the heart’s electrical activity. The technique is easily accessible and cost-effective, related to non-invasive ECG signal recording by means of skin surface electrodes connected to an electronic measurement device (electrocardiograph) [1]. The ECG bio-signal is produced by depolarization and repolarization currents flowing in different parts of the myocardium [2]. These currents are triggered by synchronized events within the cardiac cycle and define the standard patterns seen in the ECG, including the following: P-wave (atrial depolarization) initiating atrial contraction (systole); QRS-complex (ventricular depolarization) initiating ventricular systole; T-wave (ventricular repolarization) marking the beginning of ventricular relaxation; and an isoelectric line (resting membrane potentials). The position

of the electrodes on the body surface determines the angle or lead to viewing the electrical vector of the heart, which can generally be interpreted in the three-dimensional space of the frontal, horizontal, and sagittal planes [3]. The information obtained from the standard 12-lead ECG, including six limb leads (I, II, III, avR, avL, avF) and six chest leads (V1, V2, V3, V4, V5, V6), is, however, considered as a gold standard for the diagnosis of cardiovascular pathologies [4,5]. Global measurements are made on ECG waveform templates constructed for each lead from dominant complexes in the recording. Clinically useful information is mainly derived from ECG intervals and amplitudes belonging to P-, QRS-, and T-waves, defined by several fiducial points (characteristic wave peaks, onset and offset boundaries). Their measurement, also referred to in the literature as ECG delineation or segmentation, can be made from individual lead data or from mathematical combinations of simultaneously acquired individual lead data [6]. Measurement error has an important effect on the accuracy of ECG diagnostic statements [7]; therefore, standards recommend that ECG delineation systems provide their error tolerance limits obtained on precisely annotated test sets [8]. The development of accurate and robust methods for automatic ECG delineation has been a subject of continuous research.

Over the past three decades, ECG delineation methods have focused on the proper design of pre-filters and comprehensive rules for the detection of peaks and isoelectric line crossings based on empirical thresholds for wave amplitudes, slopes, and time intervals [9–19]. The low computational resource requirements of threshold-crossing algorithms have made their implementation possible in ultra-low-power microcontrollers [16] and wearable devices with embedded FPGA boards [18].

The ECG delineation task has also been tackled by machine learning (ML) techniques. Some studies undertake a simple band-pass filter design for the purposes of ECG slope enhancement and the detection of representative samples belonging to P-waves, QRS-complexes, and T-waves by means of the K-nearest neighbor rule (KNN) [20–22] and support vector machines (SVMs) [22,23]. Others explore the potential power of more sophisticated preprocessing and delineation procedures, such as an extended Kalman smoother framework followed by a differential evolution algorithm [24]; a marginalized particle filter on a non-QRS signal and sequential Bayesian detection estimation algorithm [25]; a model based on Hermite and sigmoid functions combined with piecewise polynomial interpolation for the segmentation and low-dimensional representation of individual ECG beat segments [26]; a bidirectional hidden semi-Markov model based on the probability distributions of ECG waveform duration [27]; a multiscale morphological derivative transform-based technique [28], etc.

Given that periodic P-QRS-T patterns contain waves with different frequency content, their convolution with wavelet transform (WT) mother functions extracts time scale features with different resolutions that have been shown to be feasible for ECG delineation [6]. The first WT-based ECG delineator was presented by Li et al. in 1995 [29]. A common WT approach applies ECG decomposition and subsequent reconstruction from different decomposition levels to obtain sub-signals in frequency bands representing only QRS-complexes or P- and T-waves [30–34]. A similar effect was achieved by Yochim et al. [35], who applied a continuous WT with a varying scale of the mother wavelet for the subsequent delineation of QRS-, T-, and P-waves. Another approach applied WT reconstruction for R-wave enhancement and a subsequent search within an R-wave vicinity of the QRS boundaries, further P- and T-peaks, onsets, and offsets based on time and amplitude criteria in the ECG signal [36] or in higher WT scales [6,37–39]. In some studies, WT was applied as part of more complex ECG delineation procedures, e.g., De Lannoy et al. [40] used WT for strong ECG pre-filtering followed by a hidden Markov modeling approach for ECG delineation; Sehirli et al. [41] combined WT with moving average and zero-phase filters for ECG delineation and a KNN rule for the detection of Q-, R-, and S-waves; Fu et al. [42] applied a feature, referred to as a randomly selected wavelet transform feature and a random forest classifier adapted to infer the positions of the ECG characteristic points; Ghaffari et al. [43] used additional calculations over the ECG WT representation

such as area–curve length to detect the R-peaks by variable thresholding and subsequently to identify P- and T-peaks and waves' boundaries through amplitude and slope criteria.

In the past few years, there has been a noticeable increase in projects for ECG delineation that are related to the rapid development of deep neural network (DNN) technologies. Studies classified the input raw ECG samples as belonging either to P-wave, QRS-complex, T-wave, or no-wave by using different combinations of network layers. Common model architectures are composed of stacked non-linear convolutional filters with or without combination with temporal features extracted by recurrent layers that are either long short-term memory (LSTM) [44], bidirectional LSTM (BiLSTM) [44–49], or gated recurrent units (GRUs) [50]. Some networks keep an unchanged dimensionality of the raw ECG input in hidden and output layers [43–46], while others use bottleneck architectures, which downsample the input dimensionality by pooling operations and fully connected layers with a few output units [44,51,52]. This bottleneck approach compresses the input feature map and reduces the size of the trainable parameters in hidden layers but also extracts more abstract representations of the input raw ECG signals. High-level hierarchical features encode information important to the ECG signal itself, ignoring non-essential data, such as disturbing noises. This is demonstrated in denoising autoencoders that use symmetric architectures of encoding and decoding layers to reconstruct the input to a noise-free output under strong noisy conditions with variable frequency ranges and signal-to-noise ratios [53–55]. Recently, the effective functionality of encoder–decoder transformations has also been demonstrated for ECG signal segmentation. Most such studies found in the literature use U-net architectures, where the feature sets at each level of the encoder path are concatenated to the same level in the decoder path [56–66]. Thus, the size of the feature map in the decoder path is doubled with the idea to perceive both low- and high-level hierarchical features in the network, mitigating the potential information loss caused by the pooling operation in the encoder path. Additionally, some U-net architectures are designed to add input to output feature maps by using residual (skip) connection blocks in encoder and decoder paths [62,64], given that this technique has been shown to facilitate the training of DNNs and could solve a potential problem of vanishing gradients [67]. There are also different modifications that extend the basic U-net architecture, such as the doubled U-net structure in W-net [66] and the use of multi-headed self-attention modules [63,66]. The real need for various additional connections between hidden layers in the encoder–decoder transform NN for ECG delineation has not been thoroughly investigated in the literature so far.

With respect to the input receptive field, studies on ECG delineation handle different sets of input ECG leads, most of which use only one lead [6,13–18,20,24,25,29–33,36,39–46,48–52,58–63,66] or a combination of different lead pairs [47,59,64,65] mainly founded by the largest publicly accessible ECG databases with precise delineation labels in the PhysioNet database. Fewer studies have been found to use the three vectorcardiogram leads [9,37,57] and the standard 12-lead ECG [9,10,19,21–23,35,37,38]. The majority of cited studies process long-term raw ECG input (e.g., 10 s), with the exception of [14,19,26,34,44,52,59,65], where short beat segments or average heartbeats are subjected to delineation. The latter approach supports the robust measurement of P-, QRS-, and T-onsets, offsets, and peaks that could be further involved in methods for the detection of myocardial infarction [19], reduced ejection fraction [68], arrhythmia classification [14,69], etc.

Recently, Bock et al. [70] reported considerable inter-investigator variability in the assessment of ECG time domain parameters (P-, QRS-, and T-durations, and PQ- and QT-intervals), e.g., with a Q span ranging from 39 to 99 ms. That might be considered as a factor to compromise ECG-associated diagnosis. The objectivity of computer algorithms is underlined as their main advantage, given that the ECG parameters are determined on a “blinded” basis. However, the algorithms are limited to detect ECG patterns that they have been programmed for. Therefore, the extension of the training datasets is decisive for obtaining a more generalized performance, and this need is even more binding in ECG delineators based on DNNs.

The aim of this study is to develop a DNN-based framework for 12-lead ECG delineation that can fairly compare different encoder–decoder architectures applied on clinical and standard ECG databases and noise tests. We introduce an advanced approach for the machine-based supervised learning of DNNs with large ECG databases with extensive arrhythmia records, in contrast to traditional human-based supervised learning relying on limited annotated data. Thus, different DNN architectures can be compared in terms of their capability to reproduce the performance of a commercial ML algorithm for the measurement of the P-wave, QRS-, and T-wave boundaries in 12-lead ECG representative beats. These measurements are routinely used in clinical practice to support diagnostic decisions made by cardiologists, who can rely on the same commercial ECG system as the one used for supervised learning in this study. As a result of this study, we have concluded which are the best-performing DNN architectures, based on validation against the standard practices for the evaluation of diagnostic electrocardiographs in terms of the accuracy of interval measurements on biological ECGs (P-duration, PQ-interval, QRS-duration, and QT-interval). Additionally, robustness to noise and comparison against performances of other published studies are disclosed.

## 2. Materials and Methods

### 2.1. Representative Beat and Reference Measurements

The signal averaging of repetitive PQRST patterns is a widely used technique for noise reduction [71]. Thus, the conventional diagnostic practice based on resting ECG relies on interpreting a representative beat derived by the signal averaging of multiple cardiac cycles with similar characteristics from the entire recording, rather than interpreting single beats. This is because individual beats may be influenced by artifacts or may depend on momentary conditions (such as variations related to breathing) [4]. Essential to achieving accurate diagnostic conclusions is the precise computation of the representative beats on a beat-by-beat basis. First, the combined beats should exhibit similar morphology, characterized by the highly correlated waveforms of PQRST patterns [72]. Second, it is crucial to maintain the time alignment of individual beats to accurately reproduce the amplitudes of the PQRST waveform patterns. Common practices for time alignment include methods such as maximizing correlation, identifying maximal peaks, minimizing amplitude differences, and utilizing least squares techniques [73].

In this study, the computation of the representative beats was facilitated by means of an existing ECG processing library by SCHILLER (ETM-2.6.5, Schiller AG, Baar, Switzerland) [74]. Utilizing multi-lead ECG rhythm input, the ETM performs heartbeat classification and generates ECG beats corresponding to each detected heartbeat morphology in the record. The representative beats are extracted from the input lead set within a 1.2 s window, sampled at 500 Hz, while ensuring consistent time alignment across all ECG leads. An example of ETM-based representative beats for two heartbeat classes (normal and ventricular beats) detected in a 10 s rhythm of resting 12-lead ECG is illustrated in Figure 1. When the ECG exhibits short RR-intervals, there is a possibility that the representative beat window may encompass segments of adjacent heartbeats, as illustrated for the premature ventricular beat in Figure 1 (right plot).

The ETM-based delineation of the representative beat provides fiducial points of the PQRST pattern, including P-onset, P-offset, QRS-onset, QRS-offset, and T-offset. These are named global representative beat fiducial points, which are common to all ECG leads. They are used for the calculation of global representative beat measurements covered by the standard IEC 60601-2-25:2011 [8], including four diagnostic wave durations and intervals: P-wave (P), PQ-interval (PQ), QRS-complex (QRS), and QT-interval (QT), as depicted by line segments in Figure 1. The Schiller ETM library is validated in Kligfield et al. [75].



**Figure 1.** An illustration of the reference annotation process managed by the ECG Treatment Module (ETM, Schiller AG, Switzerland). **On top:** A 10 s rhythm of a 12-lead resting ECG record (HR00255 in PTB-XL database with original diagnosis “Incomplete bundle branch block”) and beat markers identifying two types of detected beats—N (normal sinus beat) and V (premature ventricular beat). **On the bottom:** The 12-lead ECG representative beats (duration of 1.2 s) computed for the detected two beat types—N (left) and V (right). The reference annotations for the presence of a P-wave (P-wave On/Off), fiducial points and durations of the P-wave (P, green), PQ-interval (PQ, yellow), QRS-complex (QRS, red), and QT-interval (QT, blue) are measured by the ETM.

## 2.2. Training Setting

### 2.2.1. Training Database

The learning phase of this study uses the PhysioNet PTB-XL ECG database from the Physikalisch-Technische Bundesanstalt, version 1.0.1 [76,77], which is one of the largest freely accessible clinical 12-lead ECG waveform datasets, comprising 21,837 records from 18,885 patients. The twelve standard ECG leads (I, II, III, aVL, aVR, aVF, V1–V6) were recorded at rest for 10 s into a binary format with 16-bit precision at a resolution of 1  $\mu\text{V}/\text{LSB}$

and sampling frequency of 500 Hz. The database was released with the primary purpose for the evaluation of machine learning algorithms, making certain that the whole database is a rich representation of healthy controls and pathologic ECG rhythms. This study considers all records without stratification to the diagnostic label. Nevertheless, it is worth noting that the reliability of the training results was justified with the variety of clinical rhythms available, including annotations for sinus rhythm (16,782 recordings), sinus bradycardia (637), atrial fibrillation (1514), atrial flutter (73), sinus tachycardia (826), sinus arrhythmia (772), supraventricular arrhythmia (157), supraventricular tachycardia (27), paroxysmal supraventricular tachycardia (24), normal functioning artificial pacemaker (296), bigeminal (82) and trigeminal pattern (20), ventricular premature complex PVC (1146), premature complex (10), and atrial premature complex PAC (398), where each recording could have one or several diagnoses [77].

The ETM was used for the extraction of representative beats and reference measurements from the PTB-XL ECG database, according to Section 2.1. Multiple representative beats may be extracted from one recording to represent all beat types detected by the ETM within the original 10 s rhythm, as demonstrated in the example in Figure 1. This includes both normal and ventricular ectopic beats observed in a single recording. A total of 24,232 representative beats were extracted, from which 20,955 beats (86.5%) were identified with a P-wave, and 3277 beats (13.5%) were identified without a P-wave (such as atrial fibrillation beats, ventricular beats, paced beats, nodal beats, etc.).

### 2.2.2. Training Phase I: Delineation Model

ECG delineation is performed using a DNN model, which takes as input raw data samples from a 12-lead representative beat and produces probabilities for each input sample, indicating its association with one of three fundamental ECG waves/intervals: P-wave, QRS-complex, and QT-interval.

The input data are configured as a tensor with size ( $W \times \text{Ch\_in}$ ), where

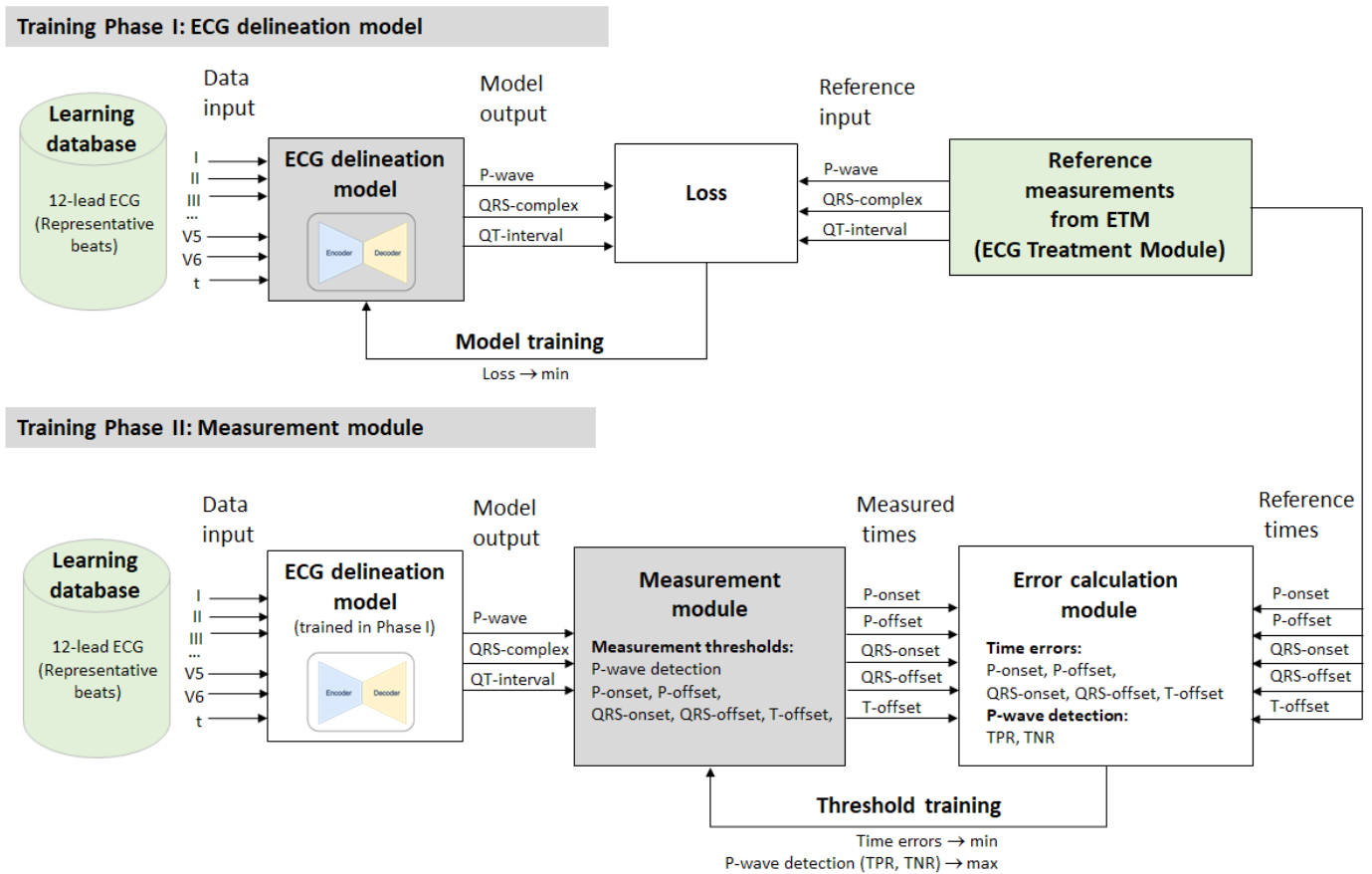
- $W = 512$  represents the analysis window (samples), including a representative beat with a duration of 1.024 s (sampling frequency of 500 Hz).
- $\text{Ch\_in} = 13$  represents the number of input channels, including 12 channels for standard 12 ECG leads (I, II, III, aVR, aVL, aVF, V1–V6) and 1 channel for a linear time vector ( $t = 1, 2, \dots, 512$ ).

Our measurement setting uses an analysis window with a central representative beat and related global annotations. However, waves of preceding and/or following beats with RR-intervals shorter than 1 s may also appear in the analysis window. Such a phenomenon is common in rapid heart rates or premature ectopic beats (see the representative beat V in Figure 1, bottom). By consent, side heartbeats are not a subject of measurement and annotation in the representative heartbeat analysis. Therefore, the time vector channel contains information that can be useful to guide the training of the DNN model on the location of the waves that need delineation. To prevent overtraining on a specific QRS-detector and representative beat synchronization, we utilize a random time offset generator. This tool reads representative beats with varying time shifts up to  $\pm 100$  ms.

The training of the DNN model for ECG delineation, also named Phase I training, is depicted in Figure 2 (top). It is supervised by the reference measurements from the ETM diagnostic module in the learning PTB-XL dataset. The ETM provides the global representative beat fiducial points of P-onset, P-offset, QRS-onset, QRS-offset, and T-offset, but these are not directly used as the reference input to the model trained for regression. Instead, we found that the most effective training is achieved by solving a binary segmentation problem for each of the following three segments: P-wave, QRS-complex, and QT-interval. As a result, the output of our DNNs is a tensor of size ( $W \times 3$ ). As a loss function for the training, we use the binary cross-entropy loss:

$$\text{Loss} = -\frac{1}{W} \sum_{i=1}^W (x_i \cdot \log(\hat{x}_i) + (1 - x_i) \cdot \log(1 - \hat{x}_i)), \quad (1)$$

where  $W$  is the number of input samples per representative beat,  $x_i$  is the reference binary input, and  $\hat{x}_i$  is the predicted output.

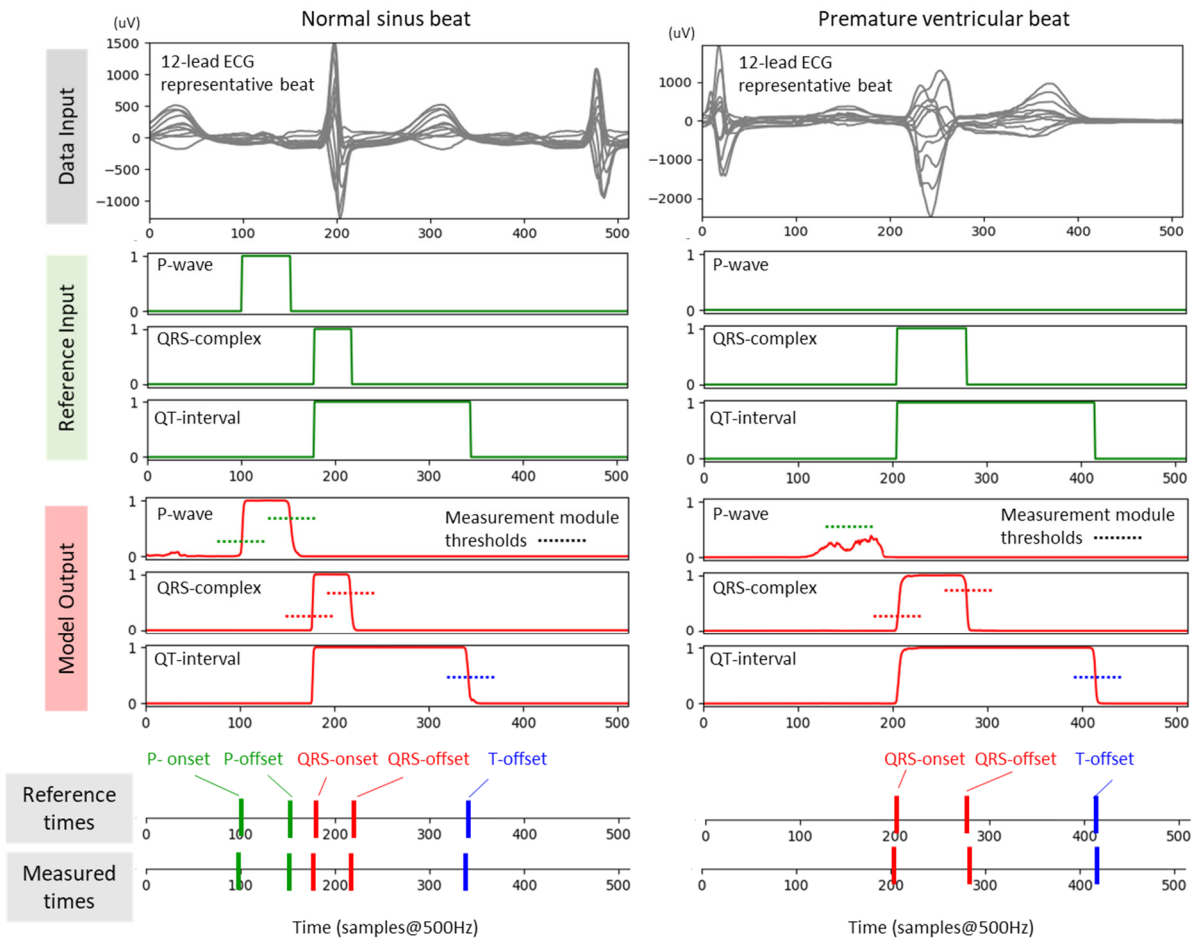


**Figure 2.** The training concept for the delineation of representative beats in 12-lead ECG, including the following: the Phase I training of the ECG delineation model (**top**) and Phase II training of the threshold-based Measurement module (**bottom**). ETM: ECG Treatment Module (Schiller AG, Switzerland); TPR: true positive rate; TNR: true negative rate.

The three target signals for the training are illustrated in Figure 3 (green traces), representing binary series [0, 1] with ones present throughout the duration of the segments of interest. The DNN outputs a probabilistic response, as shown in Figure 3 (red traces).

### 2.2.3. Training Phase II: Measurement Module

In Phase II of the training, we address the challenge of interpreting the probabilistic outputs of the ECG delineation model. As shown in Figure 3, these outputs exhibit limited steepness on the rising and falling edges, leading to uncertainty in identifying the fiducial points of the waves' onsets and offsets. According to the training scheme in Figure 2 (bottom), we use a Measurement module as a postprocessing step to the DNN to precisely measure the times of the P-onset, P-offset, QRS-onset, QRS-offset, and T-offset. They are identified by a threshold-crossing method that uses five thresholds, according to the illustration in Figure 3 (left): two thresholds for the rising and falling edges of the P-wave output (P-onset, P-offset); two related thresholds for the QRS-complex output (QRS-onset, QRS-offset); and one threshold for the falling edge of the QT-interval output (T-offset).



**Figure 3.** An illustration of the data used in the training phase in Figure 2, including two representative beats: normal sinus beat (recording HR12720 in PTB-XL database) and premature ventricular beat (recording HR11445 in PTB-XL database). The time scale (x-axis) in all plots corresponds to the size of the data frames as defined by the ECG sampling frequency of 500 Hz and representative beat duration of 512 samples (1.024 s). Data input (12 gray traces, overlapped): the representative beat seen in 12 standard ECG leads (I, II, III, aVR, aVL, aVF, V1-V6). Reference input (3 green traces): P-wave, QRS-complex, and QT-interval binary signals generated by the ETM (ECG Treatment Module) and used for model training in ECG delineation. Model output (3 red traces): the probabilistic response [0;1] for the presence of the P-wave, QRS-complex, QT-interval. Reference times (vertical lines): instants of P-onset, P-offset, QRS-onset, QRS-offset, and T-offset, determined at respective transition moments of the three reference inputs. Measured times (vertical lines): instants of P-onset, P-offset, QRS-onset, QRS-offset, and T-offset, determined at moments when the three outputs of the ECG delineation model cross the respective Measurement module thresholds.

In Phase II, each threshold is trained to ensure that the fiducial point it measures at a given time ( $T_{\text{meas}}$ ) closely matches the reference time ( $T_{\text{ref}}$ ) provided by the ETM diagnostic module. The degree of alignment is assessed by the error calculation module depicted in Figure 2, which computes the time error (TE) in a representative beat ( $n$ )

$$TE(n) = T_{\text{meas}}(n) - T_{\text{ref}}(n), \quad (2)$$

and further estimates the mean and the standard deviation of the time error for a number of  $N$  representative beats in the training database:

$$\text{mean}(TE) = \frac{\sum_{n=1}^N TE(n)}{N}, \text{std}(TE) = \sqrt{\frac{\sum_{n=1}^N (TE(n) - \text{mean}(TE))^2}{N - 1}} \quad (3)$$



The Phase II training process independently adjusts each of the five thresholds, ensuring a zero mean time error ( $\text{mean}(\text{TE}) \rightarrow 0$ ) for P-onset, P-offset, QRS-onset, QRS-offset, and T-offset, respectively. Although a low TE standard deviation is also a goal ( $\text{std}(\text{TE}) \rightarrow \text{min}$ ), it cannot be essentially controlled, but rather,  $\text{std}(\text{TE})$  is reported as a measure of the variance in or reliability of the measurement process. In order to eliminate the influence of outliers or data points on the tails that may unfairly affect the arithmetic mean, Equation (3) is estimated for N beats representative to the 95th percentile of TE in the training dataset. This approach is in line with the recommendations for reporting steady working automatic algorithms for QT-interval measurement in [78].

An additional P-wave detection threshold is defined, considering a positive P-wave detection in case the DNN output equals or exceeds the threshold and a P-wave rejection in case the DNN output does not reach the threshold. The threshold is adjusted by the Receiver Operating Characteristic (ROC) curve on the total training dataset to provide maximal balanced performance between the true positive rate (TPR) and the true negative rate (TNR):

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \text{TPR} + \text{TNR} \rightarrow \text{max}, \quad (4)$$

where TP (true positives) and FN (false negatives) are the detected and rejected P-waves, respectively, in representative beats with P-waves; TN (true negatives) and FP (false positives) are the rejected and detected P-waves, respectively, in representative beats without P-waves. Note that the evaluation of P-wave detection performance relies on the TPR and TNR. To ensure that P-wave onset/offset measurements are not biased by wrong P-wave detections, TE (P-onset) and TE (P-offset) are calculated using Equation (3) exclusively for true positive P-waves.

### 2.3. Test Setting

#### 2.3.1. Test Database Used with Reference ETM Measurements

We used a public test database from the PhysioNet database, specifically the PTB diagnostic ECG database. Since the initial release of 549 resting 15-lead ECG recordings (standard 12-leads and 3 Frank leads) from 290 subjects by the Physikalisch-Technische Bundesanstalt in 2004 [79,80], this database has been widely acknowledged in ECG signal processing research and algorithmic benchmarking. It stands out for the unique representation of healthy controls and a variety of pathologies, including myocardial infarctions, arrhythmias, heart blocks, myocardial hypertrophy, etc. Notably, this database does not overlap or share records with the training PTB-XL dataset [77].

The test phase of this study used a subset of 516 ECG records (male: 377, female: 139) from the version of the PTB diagnostic ECG database included in the Physionet/Computing in Cardiology Challenge 2020 [81]. The ETM was used for the extraction of 12-lead ECG representative beats and reference measurements according to Section 2.1. More than one representative beat might be extracted from one recording, given that they are characteristic to all beat types detected by the ETM in the original 10 s rhythm. A total of 603 representative beats were extracted, from which 527 beats (87.4%) were identified with a P-wave, and 76 beats (12.6%) were identified without a P-wave (such as atrial fibrillation beats, ventricular beats, paced beats, nodal beats, etc.). They were used for the computation of the P-wave detection performance by Equation (4).

The fidelity of measurements for P-onset, P-offset, QRS-onset, QRS-offset, and T-offset is checked against ETM annotations in 603 representative beats from the test set, applying Equations (2) and (3). A non-outlier range of  $\text{mean} \pm \text{three times the standard deviation}$  is considered, assuming that the vast majority of cases (approximately 99.7%) fall within it in a normally distributed dataset [82]. Any cases lying outside this range are deemed extreme outliers, representing data points significantly divergent from the rest of the dataset, potentially skewing the analysis or affecting the validity of statistical measures. To mitigate their impact, outliers are capped to the non-outlier range, ensuring that their extreme values do not disproportionately influence the analysis. Consequently,

we provide statistical measurements for all data points in the test database but do not apply more aggressive outlier removal techniques that result in data loss by dropping part of the observations.

### 2.3.2. The Test Database Used According to the Standard-Based Evaluation Practice

Recommendations for accuracy reports of automatic ECG delineation algorithms specify maximum tolerances for acceptable measurement errors of four diagnostic wave durations and intervals, as illustrated in Figure 1 (P-duration, PQ-interval, QRS-duration, QT-interval). The corresponding duration errors (DEs) are computed as the difference between the automatic duration measurement ( $D_{\text{meas}}$ ) and the reference duration value ( $D_{\text{ref}}$ ), taken from the annotations of a particular record ( $n$ ) in the test database:

$$DE(n) = D_{\text{meas}}(n) - D_{\text{ref}}(n) \quad (5)$$

The accuracy metrics report the duration error mean value ( $\text{mean}(DE)$ ) and standard deviation ( $\text{std}(DE)$ ), computed for a set of  $N$  records in the test database, using the same Equation (3) as for TE.

According to the IEC 60601-2-25:2011 standard [8], the duration errors shall be evaluated for a set of  $N = 100$  ECGs with the following sequential numbering in the Common Standards for Electrocardiography (CSE) database [83,84], including both physiological and pathological records with diverse ECG abnormalities: “MO1\_” series {001–005, 007–009, 011–017, 019, 021, 022, 024–044, 046–049, 051, 053, 055, 058–066, 068, 069, 071–088, 090, 091, 095–099, 101–108, 110, 112–116, 118, 123–125}. In order to reject the influence of outliers in the record set  $n = \{1, \dots, 100\}$ , the eight largest deviations from the mean  $|DE(n) - \text{mean}(DE)|$  shall be excluded, and the mean and standard deviation of the remaining differences shall not exceed the specified tolerances: P-wave DE (mean  $\leq \pm 10$  ms, std  $\leq \pm 15$  ms), PQ-interval DE (mean  $\leq \pm 10$  ms, std  $\leq \pm 10$  ms), QRS-complex DE (mean  $\leq \pm 10$  ms, std  $\leq \pm 10$  ms), QT-interval DE (mean  $\leq \pm 25$  ms, std  $\leq \pm 30$  ms).

A useful practice covered in the former standard IEC60601-2-51:2003 [85] concerned a set of tests for the noise immunity of automatic ECG delineation algorithms. It was required that the stability of the interval and wave duration measurements must be tested with  $N = 10$  selected ECGs in the CSE database (“MO1\_” series {008, 011, 013, 014, 015, 021, 026, 027, 042, 061}) without noise and in the presence of three kinds of noises:

- A 25  $\mu\text{V}$  r.m.s. high-frequency (HF) noise;
- A 50  $\mu\text{V}$  peak-to-peak 50/60 Hz sinusoidal power-line (PL) frequency noise;
- A 1 mV peak-to-peak 0.3 Hz sinusoidal low-frequency (LF) baseline noise.

According to the standard [85], the duration error in the presence of noise ( $DE_{\text{noise}}$ ) is determined as the differences in duration measurements between noise-free ECGs and ECGs with each type of noise:

$$DE_{\text{noise}}(n) = D_{\text{noise-free}}(n) - D_{\text{noise}}(n) \quad (6)$$

Similar to the practice with noise-free records, the mean value and standard deviation of  $DE_{\text{noise}}$  shall be calculated by Equation (3) after rejecting the influence of outliers, which are the two largest deviations from the mean. Only the results are reported, with no requirements set for acceptable limits.

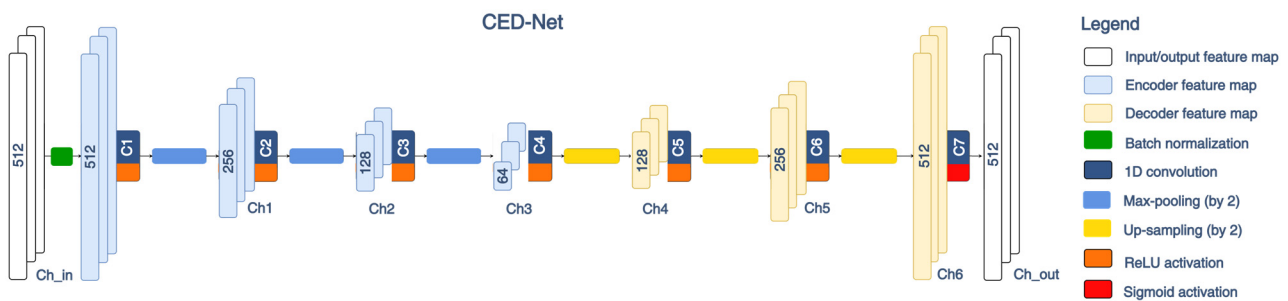
The described standard-based evaluation practice was applied in the test phase of the deep learning algorithms in this study. One representative beat per recording was taken, based on ETM calculations for the predominant beat class. The original annotations provided by the CSE database were taken as a reference.

## 2.4. ECG Delineation Model

### 2.4.1. Convolutional Encoder–Decoder Network (CED-Net)

Encoder–decoder architectures are often used in image and signal segmentation tasks [59–63,66,86]. They consist of an encoder part that downsamples the input data to extract high-level encoded features and a decoder part that upsamples these features to generate a segmentation mask. An important design consideration for the encoder–decoder is the model symmetry in terms of the total size of the feature maps and the number of neurons in all hidden layers in both the encoder and decoder parts. Another issue is the bottleneck encoding size, which must be sufficient to represent the input feature map from an informational point of view.

This study uses an ECG delineation model based on Convolutional Encoder–Decoder Network (CED-Net) in Figure 4. As defined in Section 2.2.2, the input feature map is a two-dimensional tensor with size ( $W \times \text{Ch\_in} = 512 \times 13$ ), where  $W$  is the time axis dimension of the ECG representative beat, and  $\text{Ch\_in}$  is the number of input channels, including 12 ECG leads and one time vector ( $t = 1, 2, \dots, 512$ ). The Z-score normalization of the input data is provided with the first hidden Batch normalization layer.



**Figure 4.** The architecture of the designed Convolutional Encoder–Decoder network (CED-Net).

The CED-Net design in Figure 4 uses an encoder part with three convolutional blocks, which downsample the input feature map progressively from a size of 512 (input) to 256, 128, and 64 (bottleneck). One encoder convolutional block includes a 1D convolutional layer (C1, C2, or C3) with rectified linear (ReLU) activation, followed by a max pooling operation ( $\text{pool\_size} = 2$ ). The transformation of the encoder feature map  $x_i$  after the  $i^{\text{th}}$  convolutional block can be represented as follows:

$$x_i = \text{maxpooling}(\text{ReLU}(W_i \cdot x_{i-1} + b_i), \text{pool\_size}), \quad (7)$$

where  $(.)$  denotes 1D convolution operation;  $(W_i, b_i)$  is the matrix with weights and biases of the respective encoder convolutional layer;  $\text{ReLU} = \max(x, 0)$  is the ReLU non-linear activation;  $\text{maxpooling}(x, \text{pool\_size})$  denotes a pooling operation that calculates the maximum value for “pool\_size” temporal steps along the time axis of the feature map  $x$ .

The CED-Net decoder part mirrors the encoder by using three decoder convolutional blocks, which gradually increase the resolution of the feature maps from 64 (bottleneck) to 128, 256, and 512. One decoder convolutional block includes a 1D convolutional layer (C4, C5, or C6) with ReLU activation, followed by an upsampling operation ( $\text{size} = 2$ ). The transformation of the decoder feature map  $\hat{x}_i$  after the  $i^{\text{th}}$  convolutional block is denoted as follows:

$$\hat{x}_i = \text{upsampling}\left(\text{ReLU}\left(\widehat{W}_i \cdot \widehat{x}_{i-1} + \widehat{b}_i\right), \text{size}\right), \quad (8)$$

where  $(\widehat{W}_i, \widehat{b}_i)$  is the matrix with weights and biases of the respective decoder convolutional layer;  $\text{upsampling}(x, \text{size})$  denotes an operation that repeats each value of the feature map  $\widehat{x}_{i-1}$  “size” times.

The last convolutional layer (C7) transforms the CED-Net output according to the segmentation task defined in Section 2.2.2 and presented in Figure 3. The output feature

map is provided in a tensor with size ( $W \times \text{Ch\_out} = 512 \times 3$ ), where  $W$  represents the same time axis dimension as the input ECG representative beat;  $\text{Ch\_out}$  corresponds to the number of 3 output channels (P-wave, QRS-complex, QT-interval). The convolution C7 is followed by a sigmoid activation function to force the output  $z$  in the probability range  $p \in [0;1]$ :

$$\text{sigmoid}(z) = 1 / (1 + e^{-z}) \quad (9)$$

CED-Net trainable parameters are equal to the total number of weights and biases of all convolutional layers, which are determined by the corresponding kernel size ( $K$ ) and number of channels ( $\text{Ch}$ ):

$$\text{Trainable parameters} = \sum_{i=1}^N \text{Ch}_i (K_i \text{Ch}_{i-1} + 1), \quad (10)$$

where  $N = 7$  convolutional layers (C1–C7);  $\text{Ch}_{i=0}$  refers to the number of input channels.

The kernel size and number of hidden channels are the subject of model optimization, as further disclosed in the Section 3.

#### 2.4.2. Modifications of CED-Net Architecture with Residual and Recurrent Connections

To explore the possibility of improving the performance of the basic CED-Net architecture, we propose three modifications by adding new layers or skip connections in the encoder–decoder path, presented in Figure 5. The structure of the new CED-Net architectures is motivated by recognized techniques that can either follow long-term dependencies (LSTM layers) or improve training by combining low- and high-level hierarchical features through residual connections (U-net and Res-net). Short descriptions of the networks are further presented.

1. CED-LSTM-Net (Figure 5, top): This is a CED-Net architecture with an LSTM layer provided in the decoder feature map to learn long-term dependencies in PQRST patterns. LSTM is a recurrent neural network layer with embedded memory cells, which act as accumulators of the state information [44]. The memory cells' decision to forget or remember a certain part of the information in the input time series is regularized by input, forget, and output gates, activated by non-linear sigmoid and hyperbolic tangent functions. Our application uses one memory cell per channel, defined by the dimension of  $\text{Ch}_6$  in Figure 5.
2. CED-U-Net (Figure 5, middle): This is a CED-Net architecture with residual connections between symmetrical encoder and decoder feature maps, typical for U-nets [87]. Thus, the convolution is applied on concatenated information from low- and high-order feature maps:

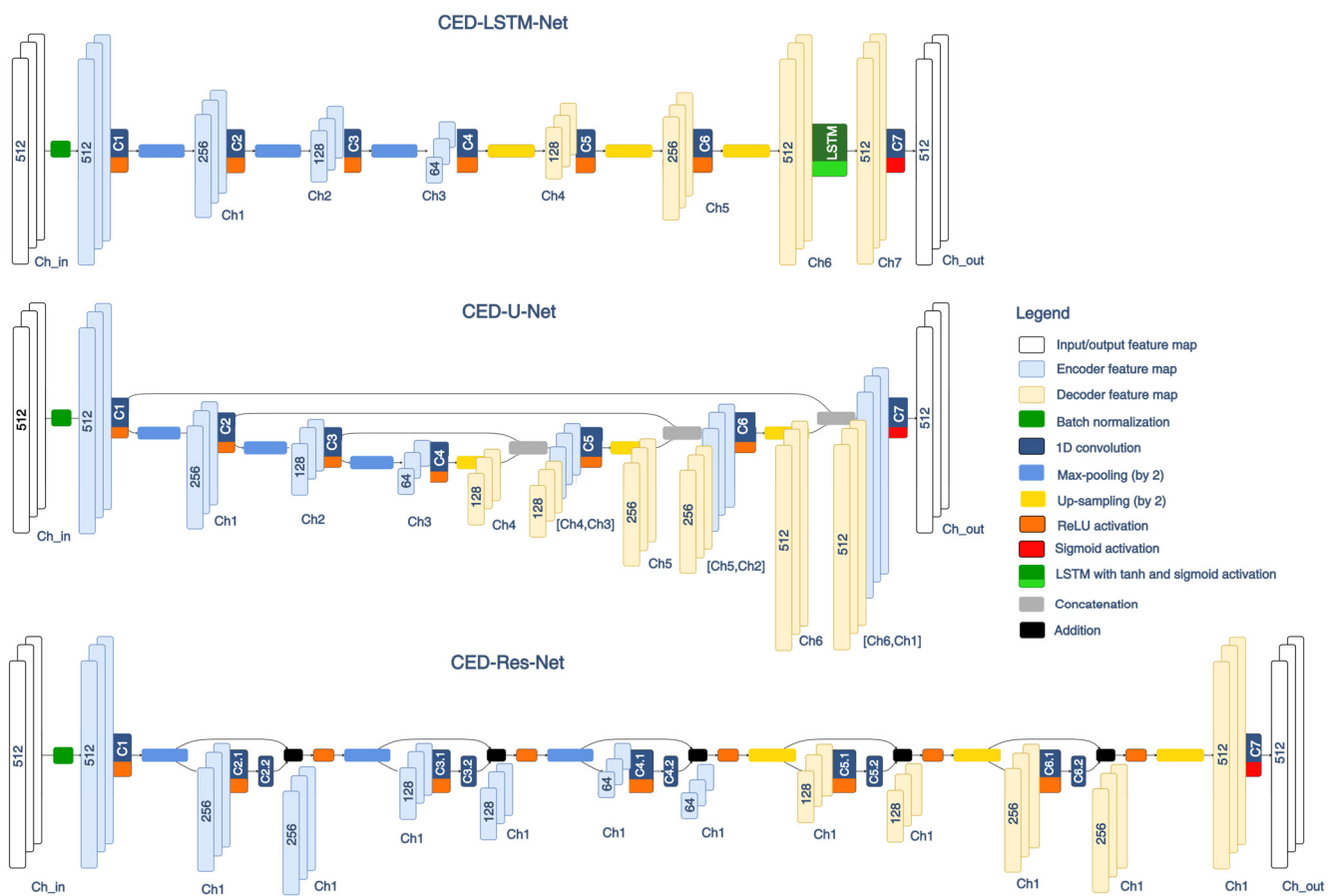
$$D_{i+1} = \text{ReLU}(W \cdot [D_i, E_j] + b), \quad (11)$$

where  $[D_i, E_j]$  means the concatenation of symmetrical decoder and encoder feature maps, denoted as  $[\text{Ch}_4, \text{Ch}_3]$ ,  $[\text{Ch}_5, \text{Ch}_2]$ , and  $[\text{Ch}_6, \text{Ch}_1]$  in Figure 5.

3. CED-Res-Net (Figure 5, bottom): This is a CED-Net architecture, including a sequence of five residual blocks. In the literature, a residual block takes an input  $x$  and produces an output  $F(x) + x$  by elementwise addition, followed by a non-linear activation function, usually ReLU [67]. In our design,  $F(x)$  represents a sequence of two convolutions, and the output  $y$  can be presented as follows:

$$y = \text{ReLU}(W_2 \cdot \text{ReLU}(W_1 \cdot x + b_1) + b_2 + x), \quad (12)$$

where  $(W_1, b_1)$ ,  $(W_2, b_2)$  are the matrices with weights and biases of the first and second convolution of the residual block, respectively.



**Figure 5.** Modifications of the basic CED-Net architecture in Figure 4 by adding an LSTM layer (CED-LSTM-Net), U-net residual connections (CED-U-Net), and residual blocks (CED-Res-Net).

### 3. Results

This section presents different results obtained in the training, optimization, and test phases of the developed CED-Net architectures in the ECG delineation task. To facilitate the understanding of such a relatively large amount of information, we illustrate the global flow of results in Figure 6. It is organized in three subsections:

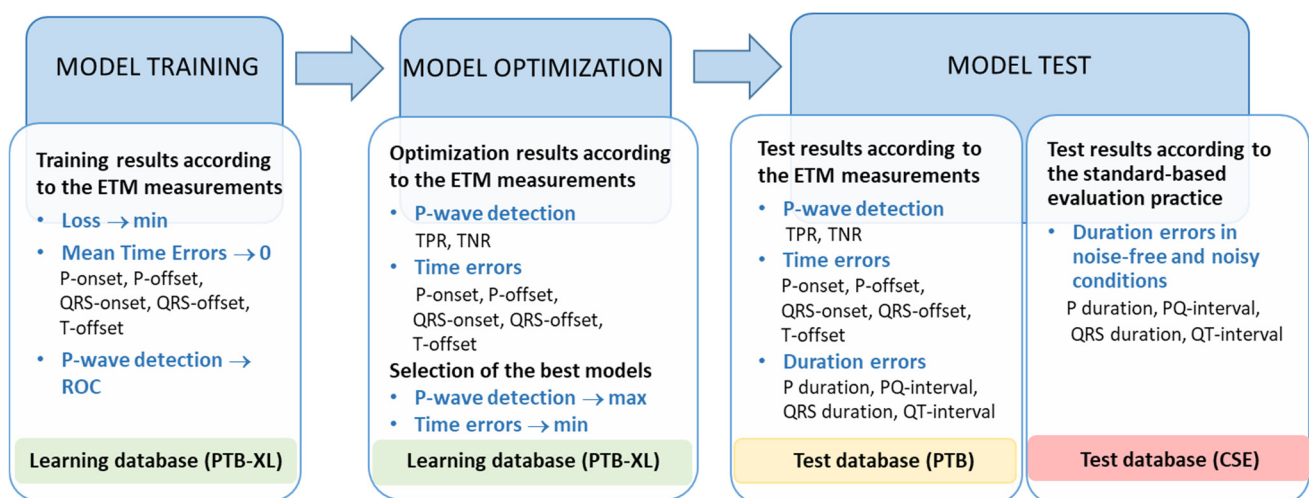
1. Model training: Shows the training results for a single model in order to illustrate the global principle of training all models.
2. Model optimization: Derives trained models according to specific hyperparameter optimization; considers a relative ranking of trained models based on their performance in the training dataset; selects the best trained models.
3. Model test: Presents the results of the best trained models with two types of tests applied on independent test datasets:
  - (1) Test results according to the ETM measurements;
  - (2) Test results according to the standard-based evaluation practice.

#### 3.1. Model Training

##### 3.1.1. Phase I Training

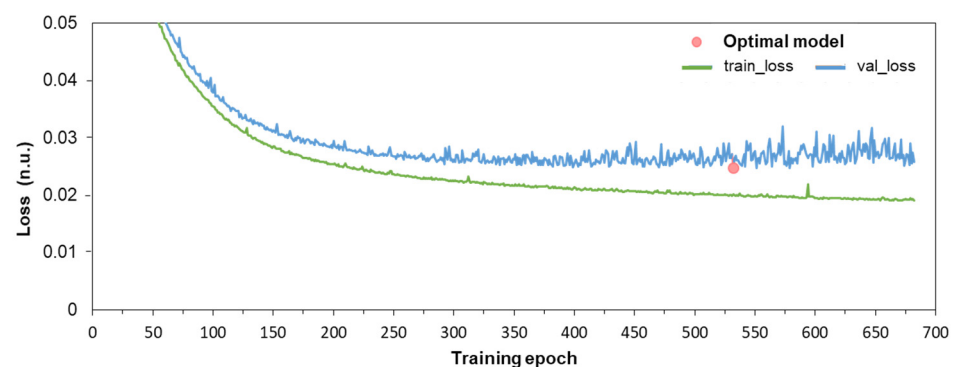
This section explains the learning strategy of a CED-Net model included in the Phase I training scheme for ECG delineation in Figure 2. The iterative training process evaluates the loss by Equation (1) in two randomly shuffled subsets of the training PTB-XL database: the training subset (70%) and validation subset (30%). These two subsets proportionally reflect the statistical characteristics of the training database, which is an essential requirement to have validation data representative of the training data. The training loss is used to

estimate how well the model fits its hidden neural weights and biases to the training data, while the validation loss estimates how well the trained model fits new data. In general, the learning curves show the trends of training and validation losses over iterative training epochs and are used to determine the optimal model fit (minimal validation loss), given that scenarios for model underfit and overfit should be avoided. Underfitting occurs either when the model is unable to accurately represent the training data (both training and validation are high) or when the validation loss decreases but has not reached a minimum. Overfitting is observed when the model performs well on training data (training loss is decreasing) but poorly on the validation set (validation loss is increasing or stagnating). Therefore, our study follows the training strategy with early stopping if the validation loss does not improve for a preset epoch tolerance. Furthermore, dropout (rate = 0.1) is applied to each convolutional layer as a regularization tool to prevent overtraining on specific data by randomly setting a fraction of the neurons to zero during training.



**Figure 6.** An illustration of the results' flow associated with the training, optimization, and testing of the developed ECG delineation models.

Figure 7 illustrates the learning curves of a CED-Net model whose training was initiated for a maximum of 1000 epochs with an early stopping tolerance of 150 epochs. Both training and validation losses are smoothly decreasing and show no findings for underfitting and overfitting. The training was stopped after 682 epochs, given that the minimum validation loss was found at epoch 532. This model was considered optimally trained and was used for further evaluation steps.

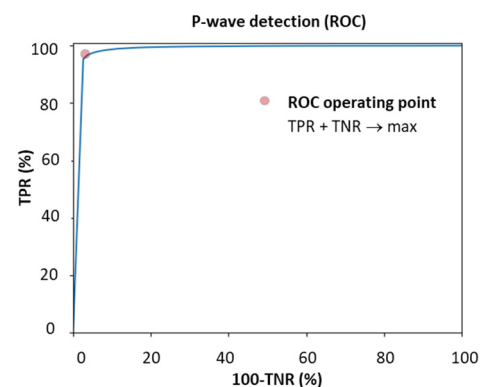


**Figure 7.** Learning curves of a CED-Net model for ECG delineation, presenting the change in the training loss (train\_loss) and validation loss (valid\_loss) over training epochs. The validation loss at epoch 532 is highlighted to represent the selected trained model with optimal fit (val\_loss  $\rightarrow$  min).

### 3.1.2. Phase II Training

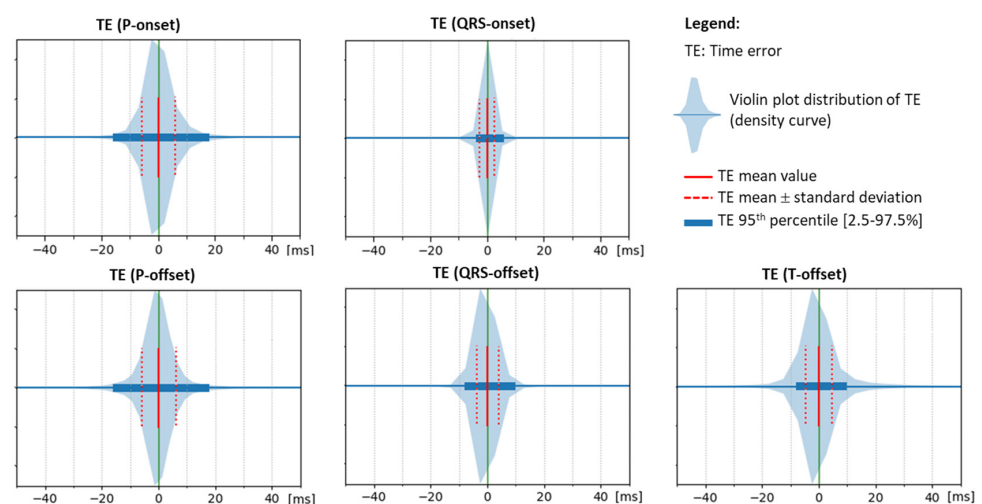
This section explains the strategy for the training of the Measurement module thresholds included in the Phase II training scheme for ECG delineation, depicted in Figure 2 and defined in Section 2.3.2.

The P-wave detection threshold is defined using the learning ROC curve in Figure 8. This ROC curve represents the relation between the TPR and TNR for P-wave detection, both calculated for different values of the threshold applied to the CED-Net output (P-wave). The operating point is selected for providing balanced performance:  $TPR + TNR \rightarrow \max$ .



**Figure 8.** ROC curve for P-wave detection, illustrating the training of the measurement threshold applied to the CED-Net output (P-wave) to ensure  $TPR + TNR \rightarrow \max$  in the training database (PTB-XL). TPR: true positive rate, TNR: true negative rate.

The five thresholds for the detection of the P-onset, P-offset, QRS-onset, QRS-offset, and T-offset in the respective CED-Net outputs are adjusted to provide a zero mean time error ( $\text{mean}(\text{TE}) \rightarrow 0$ ) in the training database. This is justified in Figure 9, which illustrates the density distributions of the corresponding time errors, showing a relatively symmetrical spread around 0. The width of TE distributions is quantified by the 95th percentile and standard deviation, considering that the lower they are, the higher the reliability of the measurement process. Deduced from the graphs, the training was the most reliable for the QRS-onset and the least reliable for the P-onset and the P-offset.

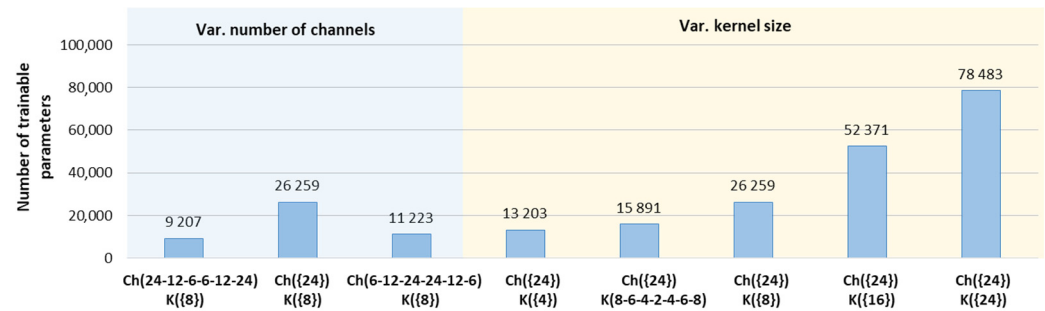


**Figure 9.** Violin plot density distributions of five time errors: TE (P-onset), TE (P-offset), TE (QRS-onset), TE (QRS-offset), and TE (T-offset), illustrating the training of the measurement thresholds applied to the CED-Net outputs (P-wave, QRS-complex, QT-interval) to ensure mean time errors  $\rightarrow 0$  in the training database (PTB-XL). The distributions are additionally indicated by the ranges of the mean  $\pm$  standard deviation computed for the 95th percentile of TE.

### 3.2. Model Optimization

#### 3.2.1. The Number of Channels and Kernel Size of the Basic CED-Net Architecture

The basic CED-Net architecture in Figure 4 is optimized by setting variable numbers of channels (Ch1–Ch6) and kernel sizes (K1–K7) for sequential convolutional layers (C1–C7), as indicated in the x-axis labels of Figure 10.



**Figure 10.** The number of trainable parameters of CED-Net models generated to optimize the hyperparameters of convolutional layers in the architecture of Figure 4. On the left (blue section): models with a variable number of channels (Ch1 to Ch6). On the right (yellow section): models with a variable kernel size (K) of seven 1D convolutions (C1 to C7). In the x-label, the values in brackets apply to all layers, specified as a series of numbers or as a single numeric constant for all layers.

The first-level CED-Net optimization uses a fixed kernel size  $K(\{8\})$  for all convolutions in Figure 4 and explores three strategies for setting variable channels:

- $Ch(\{24\})$  denotes a fixed number of 24 channels for Ch1–Ch6, selected to double the number of ECG leads to prevent information loss;
- $Ch(24-12-6-6-12-24)$  and  $Ch(6-12-24-24-12-6)$  respectively indicate a decreasing and increasing number of channels from bottom to top of the encoder and mirroring them in the decoder.

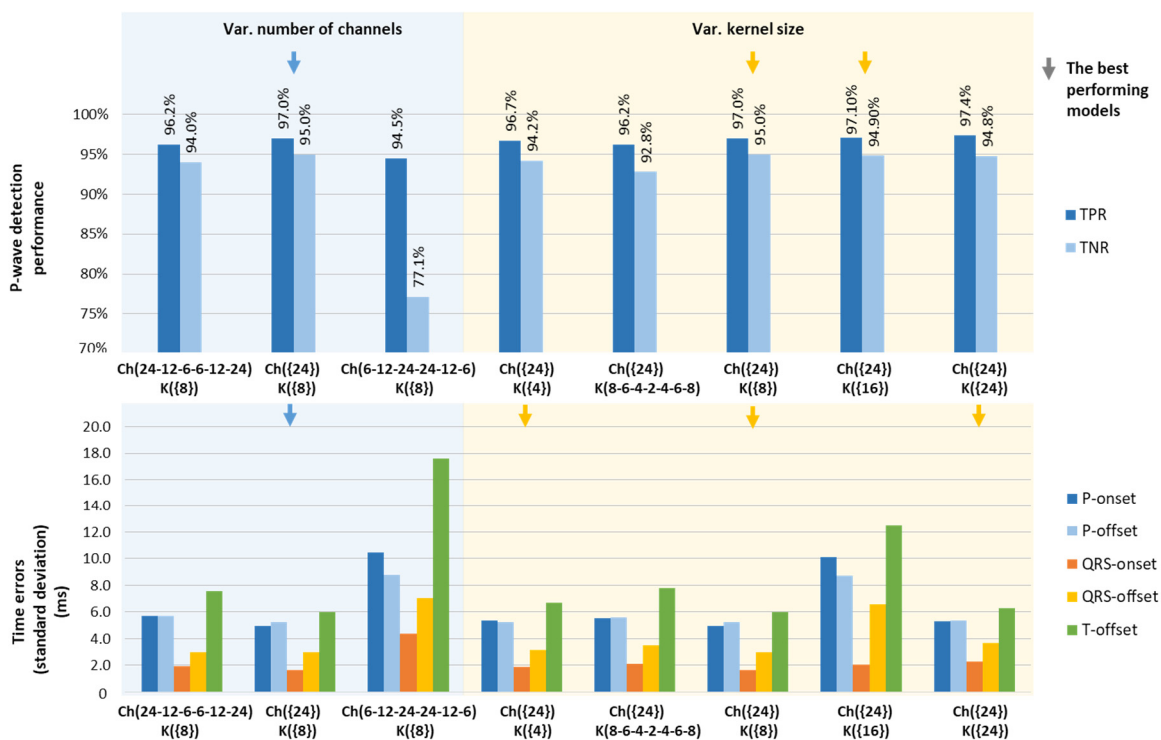
The numbers of trainable parameters in Figure 10 (blue section) indicate that increasing and decreasing numbers of channels create models that are, respectively, from 2.3 to 2.9 times lighter than the redundant model with fixed 24 channels. Nevertheless, our optimization strategy selects the best-performing model  $Ch(\{24\})$ , which is highlighted in Figure 11 (blue arrow), presenting maximal P-wave detection performance (TPR = 97%, TNR = 95%) and  $\text{std}(TE) \rightarrow \min$  (P-onset = 4.9 ms, P-offset = 5.3 ms, QRS-onset = 1.7 ms, QRS-offset = 3 ms, T-offset = 6 ms).

The second-level CED-Net optimization uses the highlighted above optimal number of channels  $Ch(\{24\})$  and explores various kernel sizes:

- $K(\{4\})$ ,  $K(\{8\})$ ,  $K(\{16\})$ , and  $K(\{24\})$  denote a fixed number of 4, 8, 16, and 24 kernels for the seven 1D convolutions (C1–C7) shown in Figure 4;
- $K(8-6-4-2-4-6-8)$  indicates a decreasing number of kernels from the bottom to the top of the encoder and mirroring them in the decoder of Figure 4.

Derived from Equation (10), the number of trainable parameters in Figure 10 (yellow section) is linearly dependent on the kernel size. Therefore, a stepwise increase in kernel size (from 4 to 24) allows for the comparison of models with gradually increasing complexity. Although the complexity range of the models is up to six times (from 13,203 to 78,483 parameters), we have not observed substantial differences in their training performance, presented in Figure 11 (yellow section). Nevertheless, the best P-wave detection, (TPR, TNR)  $\rightarrow \max$  is highlighted for two models  $K(\{8\})$ ,  $K(\{16\})$ . The minimum time error,  $\text{std}(TE) \rightarrow \min$  is marked for three models  $K(\{4\})$ ,  $K(\{8\})$ , and  $K(\{24\})$ . Thus, the  $K(\{8\})$ ,  $Ch(\{24\})$  model performed the best with both optimization criteria, being balanced in terms of training parameters, hence our final choice for the optimal CED-Net hyperparameters.





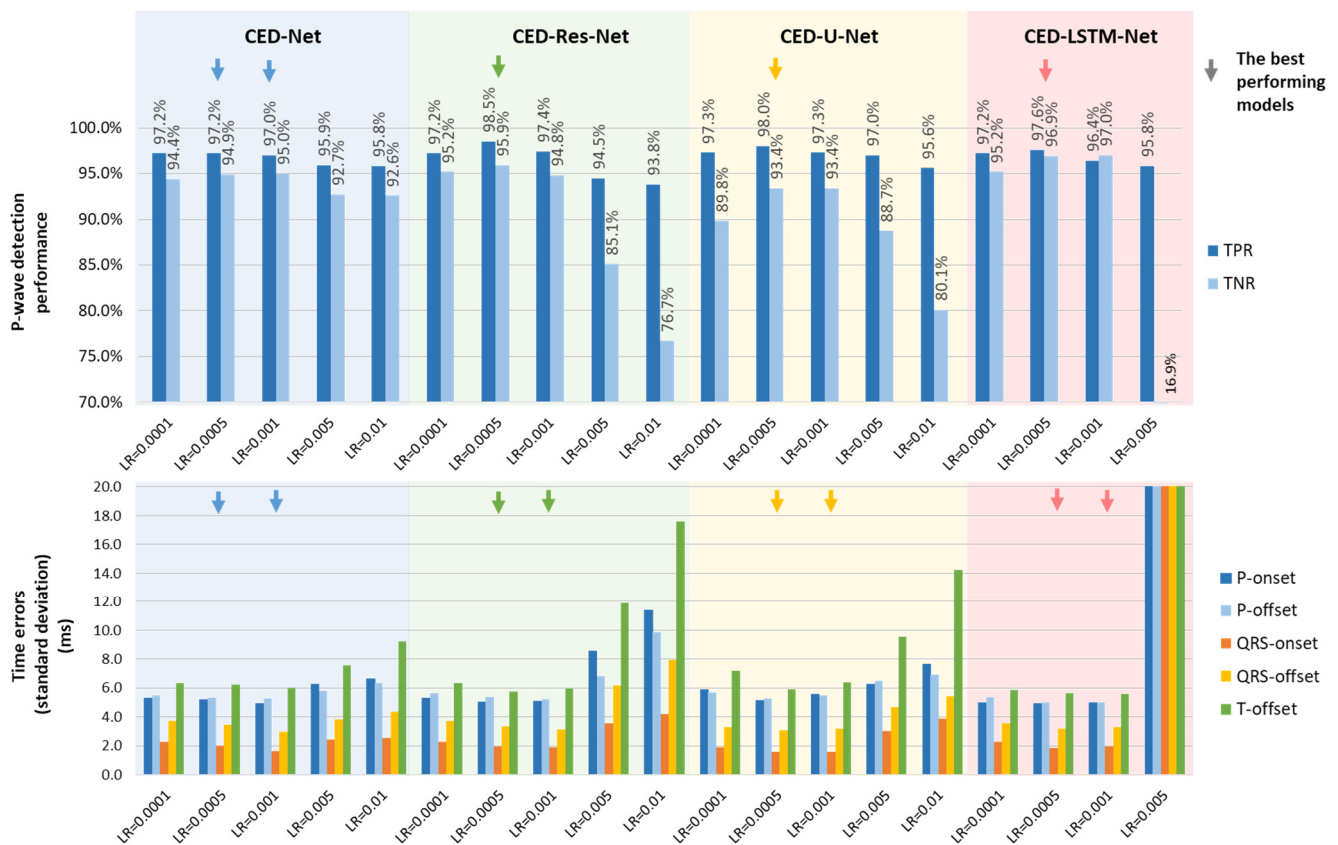
**Figure 11.** Training performance of CED-Net models subjected to hyperparameter optimization of convolutional layers, as explained Figure 10. **On top:** P-wave detection performance. **On bottom:** Time errors (standard deviations) of five fiducial points. Models with best training performances on top (TPR → max, TNR → max) and bottom (time errors → min) are marked with arrows.

### 3.2.2. Learning Rates of Different CED-Net Architectures

The CED-Net optimal hyperparameters  $K(\{8\})$ ,  $Ch(\{24\})$  are applied to the architectures in Figure 5, resulting in four models ordered by complexity (number of trainable parameters): CED-Net (26,259), CED-LSTM-Net (30,963), CED-U-Net (36,051), and CED-Res-Net (49,419). Given the different layers and connectivity, it can be assumed that models can be optimally trained with specific learning rates. Therefore, five learning rates  $LR = \{0.0001, 0.005, 0.001, 0.005, 0.01\}$  are used for the comparative study in Figure 12. Note that CED-LSTM-Net ( $LR = 0.01$ ) is not shown because it failed to train in several independent runs. The best trained models are highlighted for each architecture, noting that  $LR = 0.005$  and  $LR = 0.001$  are almost equally effective for training all models in terms of  $std(TE) \rightarrow \min$ , while  $LR = 0.005$  provides the best P-wave detection (TPR, TNR) → max. Therefore, our final choice for the best model representative from each architecture is trained with  $LR = 0.005$ .

### 3.3. The Test Results According to the Reference ETM Measurements

Applying the test concept in Section 2.3.1 and the test scheme in Figure 6, we reported an independent test evaluation of the best trained models (CED-Net, CED-Res-Net, CED-U-Net, and CED-LSTM-Net) in Table 1. All models present P-wave detection performance in the narrow range (94.7–97.9%), as well as very similar ranges for the time and duration errors with respect to mean values (−2.6 to 2.4 ms) and standard deviations (2.9 to 11.4 ms).



**Figure 12.** Training performance of CED-Net, CED-Res-Net, CED-U-Net, and CED-LSTM-Net architectures for five learning rates  $LR = \{0.0001, 0.0005, 0.001, 0.005, 0.01\}$ . **On top:** P-wave detection performance. **On bottom:** Time errors (standard deviations) of five fiducial points. Models with best training performances on top (TPR  $\rightarrow$  max, TNR  $\rightarrow$  max) and bottom (time errors  $\rightarrow$  min) are marked with arrows.

**Table 1.** Test performance of CED-Net, CED-Res-Net, CED-U-Net, and CED-LSTM-Net models with PTB diagnostic database. Categorical data for P-wave detection performance are presented as percentage (number). Continuous data for time and duration errors are presented as mean value ( $\pm$ standard deviation) in non-outlier range.

PTB Diagnostic Database	CED-Net	CED-Res-Net	CED-U-Net	CED-LSTM-Net
<b>P-wave detection</b>				
TPR, % (number)	95.6% (504/527)	97.2% (512/527)	97.9% (516/527)	96.0% (506/527)
TNR, % (number)	97.4% (74/76)	94.7% (72/76)	94.7% (72/76)	94.7% (72/76)
<b>Time Errors</b>				
P-onset, ms	$-2.3 (\pm 7.3)$	$-2.6 (\pm 6.7)$	$-1.9 (\pm 6.0)$	$-2.5 (\pm 6.2)$
P-offset, ms	$-1.4 (\pm 5.2)$	$-1.7 (\pm 5.5)$	$-2.0 (\pm 5.0)$	$-1.9 (\pm 4.9)$
QRS-onset, ms	$-1.1 (\pm 3.8)$	$-1.1 (\pm 2.9)$	$-1.3 (\pm 5.4)$	$-0.9 (\pm 3.0)$
QRS-offset, ms	$1.3 (\pm 7.9)$	$1.1 (\pm 6.7)$	$0.8 (\pm 7.2)$	$1.1 (\pm 4.7)$
T-offset, ms	$-0.1 (\pm 9.9)$	$0.3 (\pm 9.3)$	$-0.5 (\pm 8.7)$	$-1.4 (\pm 9.1)$
<b>Duration Errors</b>				
P-duration, ms	$1.0 (\pm 9.3)$	$0.9 (\pm 8.5)$	$-0.2 (\pm 8.4)$	$0.7 (\pm 8.3)$
PQ-interval, ms	$1.3 (\pm 7.5)$	$1.7 (\pm 6.8)$	$0.9 (\pm 6.6)$	$1.7 (\pm 6.6)$
QRS-duration, ms	$2.4 (\pm 9.9)$	$2.3 (\pm 7.4)$	$2.1 (\pm 10.6)$	$2.0 (\pm 5.9)$
QT-interval, ms	$1.0 (\pm 11.2)$	$1.4 (\pm 9.9)$	$0.6 (\pm 11.4)$	$-0.4 (\pm 9.9)$

### 3.4. The Test Results According to the Standard-Based Evaluation Practice

Using the test concept outlined in Section 2.3.2 and the test scheme illustrated in Figure 6, we have evaluated the performance of the best trained models (CED-Net, CED-Res-Net, CED-U-Net, and CED-LSTM-Net) in delineating diagnostic wave durations and intervals (P-duration, PQ-interval, QRS-duration, QT-interval) in accordance with the standard-based guidelines using the CSE dataset. The results are presented in Table 2 and can be directly compared with the standard tolerances for the mean value and standard deviation [8]. Although most of the measurements are within the tolerances, we have found that the CED-Res-Net, CED-U-Net, and CED-LSTM-Net models are not compliant with the standard ( $\pm 10$  ms), presenting up to 4 ms larger standard deviations for PQ-interval measurements. As an alternative, the CED-Net model meets all standard tolerances.

**Table 2.** Test performance according to standard-based evaluation practice with CSE dataset. Data are presented as mean value ( $\pm$  standard deviation).

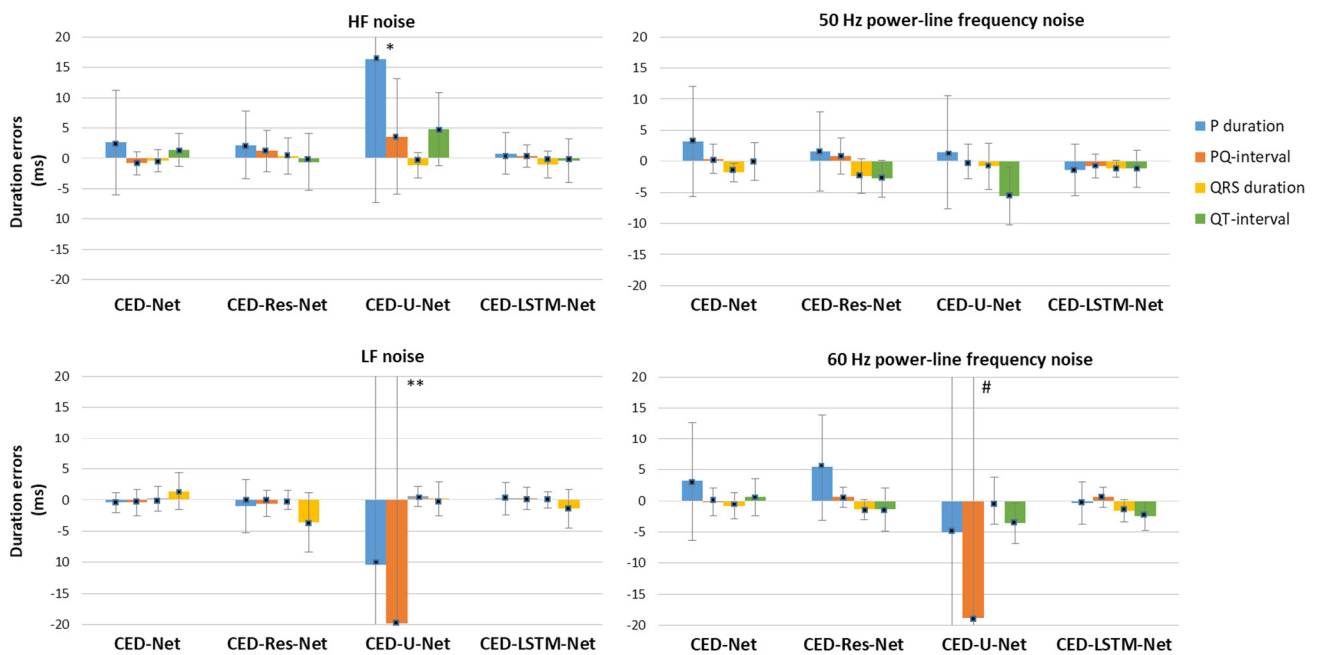
Duration Errors	Standard Tolerances [8]	CED-Net	CED-Res-Net	CED-U-Net	CED-LSTM-Net
P-duration, ms	$\pm 10$ ( $\pm 15$ )	2.6 ( $\pm 11.0$ )	$-0.6$ ( $\pm 11.5$ )	$-0.4$ ( $\pm 12.9$ )	$-0.5$ ( $\pm 11.4$ )
PQ-interval, ms	$\pm 10$ ( $\pm 10$ )	0.9 ( $\pm 5.8$ )	$-0.8$ ( $\pm 14.0$ *)	$-0.7$ ( $\pm 13.8$ *)	$-0.3$ ( $\pm 13.7$ *)
QRS-duration, ms	$\pm 10$ ( $\pm 10$ )	$-2.4$ ( $\pm 5.4$ )	$-2.3$ ( $\pm 5.3$ )	$-3.0$ ( $\pm 4.7$ )	$-2.1$ ( $\pm 5.3$ )
QT-interval, ms	$\pm 25$ ( $\pm 30$ )	$-0.7$ ( $\pm 10.3$ )	$-0.9$ ( $\pm 9.1$ )	$-0.6$ ( $\pm 11.0$ )	$-1.9$ ( $\pm 10.2$ )

\* Value not compliant with the standard tolerance.

Figure 13 compares the four network models in terms of their robustness to four types of noise: HF, LF, and PL noise at 50 Hz or 60 Hz, added to representative ECGs from the CSE database, as defined in Section 2.3.2. It is worth noting that these results can be interpreted for relative comparisons between models as no standard requirements have been set for this test in noisy conditions. According to Equation (6), the duration measured in the presence of noise is subtracted from that in the absence of noise, yielding a signed duration error. The sign indicates the effect of the noise: a positive value implies a shortened duration measurement, while a negative value implies a prolonged duration. Nevertheless, an error close to zero indicates a robust model with stable duration measurements unaffected by the presence of specific noise type. The data in Figure 13 give evidence that three of the models (CED-Net, CED-Res-Net, and CED-LSTM-Net) are robust to all types of noise, presenting a mean duration error below  $\pm 2.5$  ms that might reach up to +5 ms (prolonged P-duration) for HF and PL noises. The standard deviations of the errors are under 8 ms. In contrast to other models, CED-U-Net presents extremity errors (means up to 20 ms, standard deviations up to 65 ms), associated with the following: P-wave prolongation/shortening for HF/LF noise and QRS shortening (LF and PL 60 Hz noise). Such reduced noise immunity can be attributed to the specific U-net skip connections, which directly transfer information from input to output without passing convolutional filters.

### 3.5. Computing Resources

Deep learning on a PC requires powerful computing resources, so the experiments were conducted on a GPU-based workstation PERSY Stinger with Intel CPU Xeon Silver 4214R @ 2.4 GHz (2 processors), 96 GB RAM (Intel, Santa Clara, CA, USA), and NVIDIA RTX A5000-24 GB GPU (NVIDIA, Santa Clara, CA, USA). The ECG delineation system was running on Microsoft Windows Server 2019 Standard. All software modules were programmed in Python 3.9.5. Keras and Tensorflow 2.9.1 were used for the neural networks' implementation, training, and test.



**Figure 13.** Duration errors in presence of noise estimated for test CSE dataset. Data show differences in wave durations in absence of noise vs. four types of added noises, depicted as mean values (color bars + dots), and standard deviations (whiskers). Extremity duration errors, appearing outside the defined y-range ( $\pm 20$  ms), are observed for CED-U-Net architecture as follows: \* HF noise ( $-16.4 \pm 23.7$  ms for P-duration), \*\* LF noise ( $-10.4 \pm 37.8$  ms for P-duration,  $-19.8 \pm 64.7$  ms for PQ-interval), and # 60 Hz power-line frequency noise ( $-5 \pm 42.7$  ms for P-duration,  $-18.8 \pm 65.2$  ms for PQ-interval).

The ECG delineation system could be considered with a hybrid structure, including a deep learning stage (CED-Net, CED-Res-Net, CED-U-Net, or CED-LSTM-Net models) for the segmentation of ECG beats and a machine learning stage (Measurement module) to identify the fiducial points of the waves' onsets and offsets (Figure 2). Each stage requires computing resources for training and inference. While extensive training resources are required for the initial design time on a PC (all models required from 300 to 400 training epochs), the inference time is finally important for evaluating the real-life applicability. Therefore, we computed the inference time of the machine learning and deep learning stages of the ECG delineation system while running in the development platform. The results were computed for 100 runs of the ECG delineation system on the test dataset, and the inference times of the machine learning and deep learning stages were separately reported as the mean value  $\pm$  standard deviation (min–max range) for the delineation of one 12-lead representative beat:

- Machine learning (Measurement module):  $176 \pm 12.3$   $\mu$ s (155–203  $\mu$ s).
- Deep learning CED-Net:  $221 \pm 16.3$   $\mu$ s (201–264  $\mu$ s).
- Deep learning CED-Res-Net:  $224 \pm 16.5$   $\mu$ s (199–262  $\mu$ s).
- Deep learning CED-U-Net:  $245 \pm 16.0$   $\mu$ s (233–287  $\mu$ s).
- Deep learning CED-LSTM-Net:  $430 \pm 16.5$   $\mu$ s (408–472  $\mu$ s).

We observe comparable inference times for CNN-based networks and about twice the runtime of the network with an LSTM layer. Nevertheless, for clinical use, these inference times are acceptable, and even longer execution times can be tolerated in the worst-case scenario when the diagnostic system is not equipped with a GPU.

## 4. Discussion

### 4.1. The Main Findings of This Study

This study presents a comprehensible description of the methodological concept applicable to the training and testing of encoder–decoder DNNs in the 12-lead ECG delineation project related to the diagnostic measurements of P-wave, QRS-, and T-wave boundaries. The main questions typically addressed in the design of such projects are related to the following:

1. The reliability of the reference input and its availability in sufficient amounts of training and test data, ensuring the generalizability of the results.
2. The correctness of the DNN model design and training concept, ensuring the effective training of DNNs to produce outputs that adequately match the reference input.
3. The optimization of DNNs to compare various representative architectures and select the best-performing ones.
4. The validation of the results against independent test datasets and requirements defined in the standards.

The specific decisions involved in this ECG delineation study are discussed in the following subsections, in the context of the questions above and the results achieved.

#### 4.1.1. Training and Test Strategies

Robust deep learning requires substantial amounts of input data that represent normal and abnormal physiological conditions. The training sample size in this study includes 24,232 representative beats with various pathologies found in one of the largest public repositories with clinical 12-lead ECGs, such as the PhysioNet PTB-XL database. This sample size is in the range of the DNN trainable parameters (from about 9000 to 78,000 in Figure 10). Nevertheless, it can be considered sufficient for training small and large models that present comparable validation performances (Figure 11). This is consistent with the results of the study by Götz et al. [88], showing that deep convolutional neural networks can provide equally good classification performance when the training sample size is from 2 times higher down to 10 times lower than the trainable parameters. The generalizability of the best models is justified in inference using data from another sample pool (patients) in the PTB diagnostic database, as well as the CSE database as required in the International standard IEC 60601-2-25:2011 [8].

Given that the PhysioNet PTB-XL and PTB diagnostic databases do not contain delineation information for individual or representative beats, they are not usable in the traditional scenario with human-based supervised learning, which is usually limited to sparse annotated data. Therefore, we introduce the advanced approach for the machine-based supervised learning of DNNs, guided by the diagnostic measurements of an accurate delineation algorithm (ETM) embedded in commercial devices, which is routinely used in clinical practice to support diagnostic decisions made by cardiologists. The reference input is aligned with the traditional diagnostic output of the commercial algorithm, which is provided for representative beats, as they are assumed to be a robust representation of cardiac activity with reduced noise effects. Thus, different DNN architectures can be compared in terms of their capability to reproduce the performance of a commercial ML algorithm. It is worth noting that the presented deep learning strategy is applicable to the delineation of individual beats in case a sufficient amount of reliable reference measurements is available, although noises could have a larger effect on input data. Therefore, the noise robustness of DNN architectures is also a subject of the test results in this study. Lastly and most importantly, we ensure tests to provide compliance with standards, particularly using specific recordings from the CSE database, included in the recommendations for reporting essential performance for ECG interval measurements of diagnostic electrocardiographs.

#### 4.1.2. Model Design and Training

The primary DNN design (CED-Net in Figure 4) is based on the convolutional encoder–decoder concept, which is traditionally effective in image and signal segmentation tasks.

Specific to this study is the decision to transform 12-lead ECG input (raw data of representative beat) to three binary outputs, corresponding to the segments of interest (P-wave, QRS-complex, QT-interval). The CED-Net model itself cannot be trained for such binary output transformation but can generate three probabilistic outputs by minimizing a binary cross-entropy loss against binary reference measurements (Figure 2, top). The effective CED-Net training in such a scenario is shown in Figure 7. In the second stage, the interpretation of the CED-Net probabilistic outputs by the threshold-based Measurement module (Figure 2, bottom) is very important to the delineation performance. A comprehensive illustration of the thresholds for P-onset, P-offset, QRS-onset, QRS-offset, and T-offset, which are applied by this module, is shown in Figure 3. The training of the thresholds to provide a zero mean time error ( $\text{mean}(\text{TE}) \rightarrow 0$ ) in the training database (Figure 9) is a valid statistical approach to produce outputs that adequately match the reference input. The analysis of the P-wave detection ROC curve and the adjustment of the convex operating point (Figure 8) contribute to establishing a valid P-wave detection threshold.

#### 4.1.3. Model Optimization

In respect to DNN model optimization, basic CED-Net hyperparameters (number of channels and kernel size) are varied (Figure 10), leading to the conclusion that models can be trained with substantially comparable performances (Figure 11). Only one model failed: the one with the channel configuration Ch(6-12-24-24-12-6). This model has the smallest number of channels in the first/last layer of the encoder–decoder, clearly providing insufficient input/output feature space. As can be deduced from Figures 10 and 11, the effective strategies for providing competitive performance by resource-efficient models with a limited number of trainable parameters are as follows: (1) decreasing the number of channels from the bottom to the top of the encoder Ch(24-12-6-6-12-24), presenting a size reduced about 2.9 times in comparison to the fixed channel model Ch({24}); and (2) using a small number of kernels K({4}) as the kernel size linearly determines the model parameters. Nevertheless, our optimization strategy selected the best-performing model (Ch({24}), K({8})), which includes a sufficient number of input channels (doubling the number of input ECG leads to prevent information loss) and a suitably large, but not excessive, kernel size.

The second-level optimization considers variations in CED-Net architectures, including the recurrent LSTM layer (CED-LSTM-Net), residual connections between the encoder–decoder (CED-U-Net), and residual blocks (CED-Res-Net), comprehensively illustrated in Figure 5. Given the different layers and connectivity in the models, comparing their performance with respect to the learning rate (Figure 12) is a valuable monitor of their ability to train effectively. This ensures confidence that the final selection of the best representative for each model is optimally trained. The most robust model is CED-Net, which can find its optimum almost equally effectively at a wide LR range from 0.0001 to 0.01. In contrast, other models with residual connections presented suboptimal training at the lowest LR (0.0001) and the highest LR ( $\geq 0.005$ ). The model with an LSTM layer is most affected at a higher LR, failing to train at LR = 0.01, despite several independent training runs. These results indicate that the learning rate is an important hyperparameter for complex neural architectures and should be explored in model optimization, as presented in Figure 12.

#### 4.1.4. Model Test

Evaluating neural networks on an independent test dataset is essential in order to validate their performance, ensure that they generalize well to new data, and build confidence in their practical applicability.

The first part of the test results (Table 1) compares the capability of different DNN encoder–decoders to reproduce the measurements of the reference ML algorithm in the diagnostic PTB database. The test results (Table 1) are very similar to the training results for the highlighted best models (Figure 12), confirming the relatively similar performance of all DNN architectures. All models present P-wave detection performance with the TPR and TNR (94.7–97.9%), corresponding to an average accuracy of 96.5–96.7%, calculated as

$(\text{TPR} + \text{TNR})/2$ . None of the models are found to fail in the detection of specific fiducial points, providing the most robust detection of the QRS-onset (TE mean value of about 1 ms, standard deviation of  $\pm 2.9$ – $5.4$  ms) and the least accurate detection of the T-offset (TE standard deviation of  $\pm 8.7$ – $9.9$  ms) and the P-onset (TE mean value of about 2–3 ms). The durations of the diagnostic intervals are measured within similar error ranges, found with a mean value below 2.5 ms and a standard deviation between 5.9 and 11.4 ms. Based on the test results in Table 1, we can conclude that all DNN architectures are equally capable of reproducing the measurements of the reference ML algorithm for ECG delineation. Without any other evidence of the superiority of a given architecture, we would recommend the lightweight one, which is CED-Net with the fewest parameters.

The second part of the test results (Table 2) validates different DNNs against the standard-based evaluation practices of diagnostic electrocardiographs [8], concerning the accuracy of interval measurements on biological ECGs (P-duration, PQ-interval, QRS-duration, QT-interval). Although in the previous test, DNNs did not show essential differences with the measurements of the reference ETM algorithm embedded in commercial ECG devices, it is important to verify the generalizability of the DNN measurements with the standardized annotations in the CSE database. According to Table 2, all models present a mean duration error in a narrow range ( $-3.0$  to  $2.6$  ms), which is substantially below the most restrictive standard limit of  $\pm 10$  ms. The standard deviations of duration errors appear to be related to the type of measured ECG waves, with the lowest deviations observed for the QRS-duration ( $4.7$  to  $5.3$  ms), well within the  $\pm 10$  ms limits. For other waves, the deviations are broader, reaching up to about 14 ms. This meets the tolerances for P-duration ( $\pm 15$  ms) and QT-interval ( $\pm 30$  ms) but is critical for the PQ-interval ( $\pm 10$  ms) measured by three of the models (CED-Res-Net, CED-U-Net, and CED-LSTM-Net). Only the CED-Net model with a PQ-interval deviation of 5.8 ms meets all standard tolerances. This is a crucial practical consideration for confident ECG delineation, which can be provided by the basic encoder–decoder architecture embedded in CED-Net.

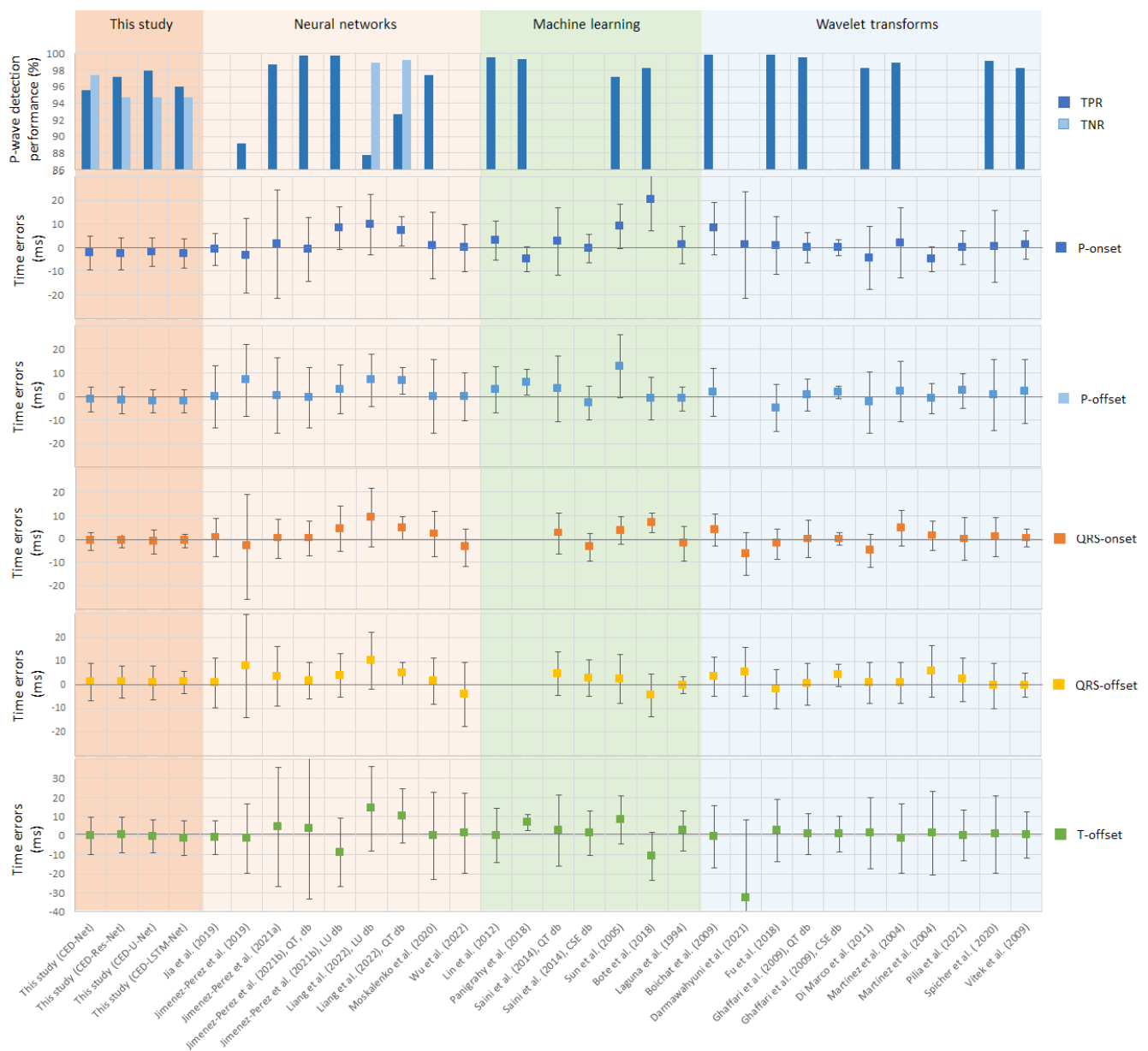
The third part of the test results (Figure 13) compares the noise immunity of DNNs in noisy conditions, specifically four types of noise: HF, LF, and PL noise at 50 Hz or 60 Hz, added to representative ECGs from the CSE database. The noise tests warn that CED-U-Net measurements might be significantly affected by all types of noise, giving extremity errors (means up to 20 ms, standard deviations up to 65 ms), associated with the following: P-wave prolongation/shortening for HF/LF noise and QRS shortening (LF and PL 60 Hz noise). This is an important finding for the reduced noise immunity of the classical U-net architecture, which is the preferable DNN design in many ECG segmentation studies [56–66]. Although the U-net skip connections have been shown to facilitate training, the direct transfer of information from input to output without passing convolutional filters is unsafe from noise immunity. The other three models (CED-Net, CED-Res-Net, and CED-LSTM-Net) are robust to all types of noise, presenting a mean duration error below  $\pm 2.5$  ms, which might reach up to +5 ms (prolonged P-duration) for HF and PL noises when compared to measurements without noise. In conclusion, we have found that encoder–decoders with convolutional filters on the signal path and without skip connections that concatenate input to output feature maps provide noise robust ECG delineation. Although these conclusions may not be important for representative beats, they are important for individual beats, where noises could have a larger effect on input data.

#### 4.2. Comparative Study with Published ECG Delineation Algorithms

##### 4.2.1. Comparison of P-Wave Detection Performances and Time Errors

The test performances of all models designed in this study (CED-Net, CED-Res-Net, CED-U-Net, and CED-LSTM-Net), as presented in Table 1, have been compared to the results reported in other published studies on ECG delineation. The comparative study is graphically summarized in Figure 14, presenting the P-wave detection performance (TPR, TNR) on top and the time errors (P-onset, P-offset, QRS-onset, QRS-offset, T-offset) on the bottom. The numerical values are reported as read from original articles (mean  $\pm$  standard

deviation, or percentages). The studies are grouped by color according to the applied methods, involving NNs or ML or WT algorithms.



**Figure 14.** Benchmarking the test performance of this study's best models (CED-Net, CED-Res-Net, CED-U-Net, and CED-LSTM-Net) against other published studies. Data are presented as percentages for the P-wave detection performance (TPR: true positive rate, TNR: true negative rate) or as mean values (markers) and standard deviations (whiskers) for the time errors for the detection of PQRST fiducial points: P-onset, P-offset, QRS-onset, QRS-offset, and T-offset. Numerical values are disclosed exactly as reported in original articles, where available. Gaps appear where data are not available. Reference studies are grouped (colored) according to the type of their methods employed for ECG delineation: neural networks (Jia et al. (2019) [65], Jimenez-Perez et al. (2019) [60], Jimenez-Perez et al. (2021a) [59], Jimenez-Perez et al. (2021b) [66], Liang et al. (2022) [63], Moskalenko et al. (2020) [58], Wu et al. (2022) [52]), machine learning (Lin et al. (2012) [25], Panigrahy et al. (2018) [24], Saini et al. (2014) [21], Sun et al. (2005) [28], Bote et al. (2018) [16], Laguna et al. (1994) [9]), wavelet transforms (Boichat et al. (2009) [33], Darmawahyuni et al. (2021) [31], Fu et al. 2018 [42], Ghaffari et al. (2009) [43], Di Marco et al. (2011) [32], Martínez et al. (2004) [6], Pilia et al. (2021) [38], Spicher et al. (2020) [34], Vitek et al. (2009) [37]).



It is worth mentioning that the comparisons in Figure 14 have been made under different conditions in respect to the databases used for reporting the results in published studies. Published NN studies used either five-fold cross-validation [52,59,65], the subject-wise splitting of a single database [58,60,63], or testing with an independent database [66] as in this study. In contrast, WT [6,31–34,37,38,43] and ML algorithms [9,16,24,25,28] did not address training data but just reported validation results on the database involved in the development process, except for WT [42] and ML [22]. The differences also concern the input data, specifically the number of ECG leads and the type of beats analyzed. This study uses 12 leads, whereas most other studies use only 2 leads. Additionally, this study focuses on representative beats, in contrast to the individual beats analyzed in other studies.

The P-wave detection performance is commonly reported by the metric TPR, which varies across studies: 89.1–99.8% for NNs, 97.2–99.5% for ML, and 98.2–99.9% for WT. The present study reports a TPR of 95.6–97.9%, falling within the range of other NNs. Since the TPR reflects the rate of false negatives, it is statistically appropriate to also report the TNR, which indicates the rate of false positives. In this study, the TNR ranges from 94.7 to 97.4%, similar to the TPR range, providing a balanced detection of the P-wave when present or absent. The average accuracy, calculated as  $(\text{TPR} + \text{TNR})/2$ , is 96.5–96.7%. This study demonstrates balanced accuracy for P-wave detection unlike [63] with a TNR up to 10% higher than the TPR (98.9–99.2% vs. 87.7–92.7%). Unfortunately, other published studies have not reported the TNR at all.

The time errors in Figure 14 can be easily interpreted: more accurate methods can be recognized by markers near zero for the mean TE and narrower whiskers representing the TE standard deviation. This study appears with mean markers in the range (−2.5 to 1.3 ms) and standard deviation whiskers in the range (2.9 to 9.9 ms) for all ECG fiducial points. This performance can be recognized in the lower range of time errors, comparable to the following studies:

- NN studies: P-onset [63], QRS-offset [63,66], and T-offset [65];
- ML studies: P-onset [9,23], P-offset [9], QRS-offset [6,9,23], and T-offset [9,23,24];
- WT studies: P-onset [37,38,43], P-offset [43], QRS-onset [37,43], QRS-offset [32,34,37,42,43], and T-offset [43].

The remaining studies report a higher mean time error and/or wider standard deviations.

#### 4.2.2. Comparison of Duration Errors Calculated According to Standard-Based Evaluation Practice

We have found only two studies [22,89] reporting duration errors of diagnostic ECG intervals according to the standard-based evaluation practice using the CSE dataset, to which this study is compared in Table 3.

- P-duration: This study exhibits one of the lowest mean P-duration errors and standard deviations similar to other studies, all within standard tolerance [8], except CSE Prog. 2 (Marquette) [89].
- PQ-interval: The standard tolerance [8] is fulfilled only by four methods, including CED-Net (this study), KNN and SVN [22], and CSE Prog. 13 (Padova) [89]. The  $\pm 10$  ms standard deviation tolerance is problematic for the other five methods, which exceed this limit by 2 to 4 ms.
- QRS-duration: All studies are well below the standard tolerance of  $\pm 10$  ms [8], with a mean value of  $\pm(0.8\text{--}3)$  ms and standard deviation  $\pm(4.7\text{--}7.3)$  ms.
- QT-interval: The standard tolerances [8] are wider, allowing for larger errors in the T-offset localization. Nevertheless, all studies do not present important problems with QT-interval measurements, having the following: mean errors within  $\pm 2$  ms (this study),  $\pm 5$  ms [22] and  $\pm 6.2$  ms [89], compared to the  $\pm 25$  ms standard tolerance; and standard deviations within  $\pm 15.4$  ms compared to the  $\pm 30$  ms standard tolerance.

In conclusion, full compliance with standard tolerances [8] can be highlighted for CED-Net (this study), KNN and SVN [22], and CSE Prog. 13 (Padova) [89].

**Table 3.** Benchmarking the duration errors of this study’s best models (CED-Net, CED-Res-Net, CED-U-Net, and CED-LSTM-Net) against other published studies according to the standard-based evaluation practice with the CSE dataset. Data are presented as the mean value ( $\pm$ standard deviation).

Study	Method	P-Duration (ms)	PQ-Interval (ms)	QRS-Duration (ms)	QT-Interval (ms)
[8]	Standard Tolerances	$\pm 10$ ( $\pm 15$ )	$\pm 10$ ( $\pm 10$ )	$\pm 10$ ( $\pm 10$ )	$\pm 25$ ( $\pm 30$ )
This study	CED-Net	2.6 ( $\pm 11.0$ )	0.9 ( $\pm 5.8$ )	−2.4 ( $\pm 5.4$ )	−0.7 ( $\pm 10.3$ )
This study	CED-Res-Net	−0.6 ( $\pm 11.5$ )	−0.8 ( $\pm 14.0$ *)	−2.3 ( $\pm 5.3$ )	−0.9 ( $\pm 9.1$ )
This study	CED-U-Net	−0.4 ( $\pm 12.9$ )	−0.7 ( $\pm 13.8$ *)	−3.0 ( $\pm 4.7$ )	−0.6 ( $\pm 11.0$ )
This study	CED-LSTM-Net	−0.5 ( $\pm 11.4$ )	−0.3 ( $\pm 13.7$ *)	−2.1 ( $\pm 5.3$ )	−1.9 ( $\pm 10.2$ )
[22]	KNN	−2.5 ( $\pm 9.3$ )	5.4 ( $\pm 9.9$ )	−2.3 ( $\pm 7.2$ )	5.0 ( $\pm 13.4$ )
[22]	SVN	−0.1 ( $\pm 7.9$ )	2.3 ( $\pm 7.1$ )	−1.1 ( $\pm 7.0$ )	3.9 ( $\pm 11.8$ )
[89]	CSE Prog. 2 (Marquette)	−12.0 * ( $\pm 17.6$ *)	−8.7 ( $\pm 12.1$ *)	−0.8 ( $\pm 7.2$ )	6.2 ( $\pm 15.4$ )
[89]	CSE Prog. 11 (Glasgow)	−2.4 ( $\pm 12.6$ )	5.1 ( $\pm 12.5$ *)	−1.4 ( $\pm 7.1$ )	3.9 ( $\pm 14.8$ )
[89]	CSE Prog. 13 (Padova)	2.8 ( $\pm 10.1$ )	−2.8 ( $\pm 8.3$ )	1.8 ( $\pm 7.3$ )	−1.1 ( $\pm 9.2$ )

\* Value not compliant with the standard tolerance.

## 5. Conclusions

This study investigates DNNs, which transform 12-lead representative beats to three diagnostic ECG intervals (P-wave, QRS-complex, and QT-interval) used for the global delineation of the representative beat (P-onset, P-offset, QRS-onset, QRS-offset, and T-offset). Although neural network-based models for ECG time series analysis are not a new frontier, to the best of our knowledge, this is the first study to directly compare the ECG delineation performances of different encoder–decoder architectures applied on clinical and standard ECG databases and noise tests. Four concepts for encoder–decoders based on a fully convolutional architecture (CED-Net) and its modifications with a recurrent layer (CED-LSTM-Net), residual connections between symmetrical encoder and decoder feature maps (CED-U-Net), and sequential residual blocks (CED-Res-Net) are involved in a fair comparison. All encoder–decoders were trained and optimized with the large PhysioNet ECG database (PTB-XL) under identical conditions, applying an advanced approach for machine-based supervised learning with a commercial ML algorithm for ECG delineation. The test results demonstrate the equal capability of all DNN architectures to reproduce the measurements of the reference ML algorithm in the diagnostic PTB database with an average P-wave detection accuracy (96.6%), time and duration error means (−2.6 to 2.4 ms), and standard deviations (2.9 to 11.4 ms). The validation according to the standard-based evaluation practices of diagnostic electrocardiographs with the CSE database outlines the CED-Net model, which measures P-duration ( $2.6 \pm 11.0$  ms), PQ-interval ( $0.9 \pm 5.8$  ms), QRS-duration ( $-2.4 \pm 5.4$  ms), and QT-interval ( $-0.7 \pm 10.3$  ms), which meet all standard tolerances. Noise tests with HF, LF, and PL noise at 50 Hz or 60 Hz confirm that CED-Net, CED-Res-Net, and CED-LSTM-Net are robust to all types of noise, mostly presenting a mean duration error below  $\pm 2.5$  ms when compared to measurements without noise. Reduced noise immunity is observed for the U-net architecture. Comparative analysis with other published studies scores this research within the lower range of time errors, highlighting its competitive performance.

In this study, we have looked at the neural networks applied to the delineation of representative beats. Future research could consider another sample pool aside from the PhysioNet ECG database, data augmentation with noise components, a detailed ECG delineation of the QRS-complex, ST-segment, and T-wave, and the delineation of multiple PQRST segments in the ECG recording rather than representative beats.

**Author Contributions:** Conceptualization, V.K., R.S. and I.J.; methodology, V.K., T.S., R.S. and I.J.; software, V.K. and T.S.; validation, V.K. and I.J.; formal analysis, V.K., T.S. and I.J.; investigation, V.K. and I.J.; resources, V.K., T.S., R.S. and I.J.; data curation, T.S. and R.S.; writing—V.K. and I.J.; writing—review and editing, V.K., T.S. and I.J.; visualization, V.K., T.S. and I.J.; supervision, R.S.; project administration, V.K.; funding acquisition, V.K. and I.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Bulgarian National Science Fund, grant number KII-06-H42/3: “Computer-aided diagnosis of cardiac arrhythmias based on machine learning and deep neural networks”.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The PTB-XL dataset is publicly available through the PhysioNet website at <https://physionet.org/content/ptb-xl/1.0.1/>, last accessed on 7 June 2024. Restrictions apply to the availability of the test CTS dataset [84], on which the authors are not permitted to release the digital data.

**Conflicts of Interest:** Author Ramun Schmid was employed by the company Schiller AG. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Krikler, D.M. Historical aspects of electrocardiography. *Cardiol. Clin.* **1987**, *5*, 349–355. [[CrossRef](#)] [[PubMed](#)]
2. Wei, X.; Yohannan, S.; Richards, J.R. Physiology, Cardiac Repolarization Dispersion and Reserve. In *StatPearls [Internet]*; StatPearls Publishing: Treasure Island, FL, USA, 2023. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK537194/> (accessed on 1 April 2024).
3. Macfarlane, P.W.; Lawrie, T.D.V. The Normal Electrocardiogram and Vectorcardiogram. In *Comprehensive Electrocardiology*; Macfarlane, P.W., van Oosterom, A., Pahlm, O., Kligfield, P., Janse, M., Camm, J., Eds.; Springer: London, UK, 2010. [[CrossRef](#)]
4. Kligfield, P.; Gettes, L.S.; Bailey, J.J.; Childers, R.; Deal, B.J.; Hancock, E.W.; Van Herpen, G.; Kors, J.A.; Macfarlane, P.; Mirvis, D.M.; et al. Recommendations for the standardization and interpretation of the electrocardiogram: Part I: The electrocardiogram and its technology: A scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: Endorsed by the International Society for Computerized Electrocardiology. *Circulation* **2007**, *115*, 1306–1324. [[CrossRef](#)] [[PubMed](#)]
5. Mason, J.W.; Hancock, E.W.; Gettes, L.S. Recommendations for the standardization and interpretation of the electrocardiogram: Part II: Electrocardiography diagnostic statement list: A scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: Endorsed by the International Society for Computerized Electrocardiology. *Circulation* **2007**, *115*, 1325–1332. [[CrossRef](#)] [[PubMed](#)]
6. Martínez, J.P.; Almeida, R.; Olmos, S.; Rocha, A.P.; Laguna, P. A wavelet-based ECG delineator: Evaluation on standard databases. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 570–581. [[CrossRef](#)] [[PubMed](#)]
7. Zywiec, C.; Celikag, D.; Joseph, G. Influence of ECG measurement accuracy on ECG diagnostic statements. *J. Electrocardiol.* **1996**, *29*, 67–72. [[CrossRef](#)] [[PubMed](#)]
8. IEC 60601-2-25:2011; Medical Electrical Equipment—Part 2–25: Particular Requirements for the Basic Safety and Essential Performance of Electrocardiographs. International Electrotechnical Commission: London, UK, 2011.
9. Laguna, P.; Jane, R.; Caminal, P. Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database. *Comput. Biomed. Res.* **1994**, *27*, 45–60. [[CrossRef](#)] [[PubMed](#)]
10. Chazal, P.; Celler, B.G. Automatic measurement of the QRS inset and offset in individual ECG leads. In Proceedings of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Amsterdam, The Netherlands, 31 October–3 November 1996.
11. Daskalov, I.; Christov, I. Electrocardiogram signal preprocessing for automatic detection of QRS boundaries. *Med. Eng. Phys.* **1999**, *21*, 37–44. [[CrossRef](#)] [[PubMed](#)]
12. Christov, I.; Simova, I. Fully automated method for QT interval measurement in ECG. *Comput. Cardiol.* **2006**, *33*, 321–324.
13. Christov, I.; Simova, I. Q-onset and T-end delineation: Assessment of the performance of an automated method with the use of a reference database. *Physiol. Meas.* **2007**, *28*, 213–221. [[CrossRef](#)] [[PubMed](#)]
14. Christov, I.; Krasteva, V.; Simova, I.; Neycheva, T.; Schmid, R. Ranking of the most reliable beat morphology and heart rate variability features for the detection of atrial fibrillation in short single-lead ECG. *Physiol. Meas.* **2018**, *39*, 094005. [[CrossRef](#)]

15. Sadhukhan, D.; Mitra, M. Detection of ECG characteristic features using slope thresholding and relative magnitude comparison. In Proceedings of the 3rd International Conference on Emerging Applications of Information Technology (EAIT), Kolkata, India, 29 November–1 December 2012; pp. 122–126. [[CrossRef](#)]
16. Bote, J.; Recas, J.; Rincon, F.; Atienza, D.; Hermida, R. A modular low-complexity ECG delineation algorithm for real-time embedded systems. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 429–441. [[CrossRef](#)] [[PubMed](#)]
17. Costa, R.; Winkert, T.; Manhaes, A.; Teixeira, J.P. QRS Peaks, P and T waves identification in ECG. *Procedia Comput. Sci.* **2021**, *181*, 957–964. [[CrossRef](#)]
18. Tang, X.; Tang, W. An ECG delineation and arrhythmia classification system using slope variation measurement by ternary second-order delta modulators for wearable ECG sensors. *IEEE Trans. Biomed. Circuits Syst.* **2021**, *15*, 1053–1065. [[CrossRef](#)] [[PubMed](#)]
19. Dohare, A.; Kumar, V.; Kumar, R. Detection of myocardial infarction in 12 lead ECG using support vector machine. *Appl. Soft Comput.* **2018**, *64*, 138–147. [[CrossRef](#)]
20. Mehta, S.S.; Shete, D.A.; Lingayat, N.S.; Chouhan, V.S. K-means algorithm for the detection and delineation of QRS-complexes in electrocardiogram. *IRBM* **2010**, *31*, 48–54. [[CrossRef](#)]
21. Saini, I.; Singh, D.; Khosla, A. K-nearest neighbour-based algorithm for P- and T-waves detection and delineation. *J. Med. Eng. Technol.* **2014**, *38*, 115–124. [[CrossRef](#)]
22. Saini, I.; Singh, D.; Khosla, A. Automated ECG delineation using machine learning algorithms. In Proceedings of the 41st International Congress on Electrocardiology, Bratislava, Slovakia, 4–7 June 2014; pp. 249–252.
23. Saini, I.; Singh, D.; Khosla, A. P- and T-wave delineation in ECG signals using support vector machine. *IETE J. Res.* **2013**, *59*, 615–623. [[CrossRef](#)]
24. Panigrahy, D.; Sahu, P.K. P and T wave detection and delineation of ECG signal using differential evolution (DE) optimization strategy. *Australas. Phys. Eng. Sci. Med.* **2018**, *41*, 225–241. [[CrossRef](#)] [[PubMed](#)]
25. Lin, C.; Giremus, A.; Mailhes, C.V.; Tourneret, J.-Y. Beat-to-beat P and T wave delineation in ECG signals using a marginalized particle filter. In Proceedings of the 20th European Signal Processing Conference (EUSIPCO), Bucharest, Romania, 27–31 August 2012; pp. 479–483.
26. Bock, C.; Kovacs, P.; Laguna, P.; Meier, J.; Huemer, M. ECG beat representation and delineation by means of variable projection. *IEEE Trans. Biomed. Eng.* **2021**, *68*, 2997–3008. [[CrossRef](#)]
27. Huo, R.; Zhang, L.; Liu, F.; Wang, Y.; Liang, Y.; Wei, S. ECG segmentation algorithm based on bidirectional hidden semi-Markov model. *Comput. Biol. Med.* **2022**, *150*, 106081. [[CrossRef](#)]
28. Sun, Y.; Chan, K.L.; Krishnan, S.M. Characteristic wave detection in ECG signal using morphological transform. *BMC Cardiovasc. Disord.* **2005**, *5*, 28. [[CrossRef](#)] [[PubMed](#)]
29. Li, C.; Zheng, C.; Tai, C. Detection of ECG characteristic points using wavelet transforms. *IEEE Trans. Biomed. Eng.* **1995**, *42*, 21–28. [[CrossRef](#)] [[PubMed](#)]
30. Chieng, T.M.; Hau, Y.W.; Omar, Z.B.; Lim, C.W. An efficient instantaneous ECG delineation algorithm. *Comput. Cardiol.* **2019**, *46*. [[CrossRef](#)]
31. Darmawahyuni, A.; Nurmaini, S.; Supriansyah, H.H.; Fauzi, M.I.R.; Rachmatullah, M.N.; Firdaus, F.; Tutuko, B. Delineation of electrocardiogram morphologies by using discrete wavelet transforms. *Indones. J. Electr. Eng. Comput. Sci.* **2021**, *22*, 159–167. [[CrossRef](#)]
32. Di Marco, L.Y.; Chiari, L. A wavelet-based ECG delineation algorithm for 32-bit integer online processing. *Biomed. Eng. OnLine* **2011**, *10*, 23. [[CrossRef](#)] [[PubMed](#)]
33. Boichat, N.; Khaled, N.; Rincon, F.; Atienza, D. Wavelet-based ECG delineation on a wearable embedded sensor platform. In Proceedings of the Sixth International Workshop on Wearable and Implantable Body Sensor Networks, Berkeley, CA, USA, 3–5 June 2009. [[CrossRef](#)]
34. Spicher, N.; Kukuk, M. Delineation of electrocardiograms using multiscale parameter estimation. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2216–2229. [[CrossRef](#)] [[PubMed](#)]
35. Yochum, M.; Renaud, C.; Jacquir, S. Automatic detection of P, QRS and T patterns in 12 leads ECG signal based on CWT. *Biomed. Signal Process. Control* **2016**, *25*, 46–52. [[CrossRef](#)]
36. Londhe, N.D.; Gupta, S.; Dewangan, D. A complete ECG signal delineation method based on Wavelet Transform. *Int. J. Biomed. Eng. Technol.* **2012**, *9*, 147–162. [[CrossRef](#)]
37. Vitek, M.; Hruběš, J.; Kozumplík, J. A Wavelet-based ECG delineation in multilead ECG signals: Evaluation on the CSE database. *IFMBE Proc.* **2009**, *25/IV*, 177–180. [[CrossRef](#)]
38. Pilia, N.; Nagel, C.; Lenis, G.; Becker, S.; Dössel, O.; Loewe, A. ECGdeli—An open source ECG delineation toolbox for MATLAB. *SoftwareX* **2021**, *13*, 100639. [[CrossRef](#)]
39. Banerjee, S.; Gupta, R.; Mitra, M. Delineation of ECG characteristic features using multiresolution wavelet analysis method. *Measurement* **2012**, *45*, 474–487. [[CrossRef](#)]
40. De Lannoy, G.; Frenay, B.; Verleysen, M.; Delbeke, J. Supervised ECG delineation using the wavelet transform and Hidden Markov Models. In Proceedings of the 4th European Conference of the International Federation for Medical and Biological Engineering, Antwerp, Belgium, 23–27 November 2008; pp. 22–25. [[CrossRef](#)]

41. Sehirlir, E.; Turan, M.K. A novel method for segmentation of QRS complex on ECG signals and classification of cardiovascular diseases via a hybrid model based on machine learning. *Int. J. Intell. Syst. Appl. Eng.* **2021**, *9*, 12–21. [[CrossRef](#)]
42. Fu, D.; Xia, Z.; Gao, P.; Wang, H.; Lin, J.; Sun, L. ECG Delineation with randomly selected wavelet feature and random forest classifier. *IEICE Trans. Inf. Syst.* **2018**, *101*, 2082–2091. [[CrossRef](#)]
43. Ghaffari, A.; Homaeinezhad, M.R.; Akraminia, M.; Atarod, M.; Daevaeiha, M. A robust wavelet-based multi-lead electrocardiogram delineation algorithm. *Med. Eng. Phys.* **2009**, *31*, 1219–1227. [[CrossRef](#)] [[PubMed](#)]
44. Peimankar, A.; Puthusserypady, S. An ensemble of deep recurrent neural networks for P-wave detection in electrocardiogram. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019), Brighton, UK, 12–17 May 2019; pp. 1284–1288. [[CrossRef](#)]
45. Peimankar, A.; Puthusserypady, S. DENS-ECG: A deep learning approach for ECG signal delineation. *Expert Syst. Appl.* **2021**, *165*, 113911. [[CrossRef](#)]
46. Abrishami, H.; Han, C.; Zhou, X.; Campbell, M.; Czosek, R. Supervised ECG Interval Segmentation Using LSTM Neural Network. In Proceedings of the International Conference Bioinformatics and Computational Biology BIOCOMP'18, Amsterdam, The Netherlands, 16–18 May 2018; pp. 71–77.
47. Malali, A.; Hiriyannaiah, S.; Siddesh, G.M.; Srinivasa, K.G.; Sanjay, N.T. Supervised ECG wave segmentation using convolutional LSTM. *ICT Express* **2020**, *6*, 166–169. [[CrossRef](#)]
48. Tutuko, B.; Rachmatullah, M.N.; Darmawahyuni, A.; Nurmaini, S.; Tondas, A.E.; Passarella, R.; Partan, R.U.; Rifai, A.; Sapitri, A.I.; Firdaus, F. Short single-lead ECG signal delineation-based deep learning: Implementation in automatic atrial fibrillation identification. *Sensors* **2022**, *22*, 2329. [[CrossRef](#)] [[PubMed](#)]
49. Nurmaini, S.; Darmawahyuni, A.; Rachmatullah, M.N.; Effendi, J.; Sapitri, A.I.; Firdaus, F.; Tutuko, B. Beat-to-Beat Electrocardiogram Waveform Classification Based on a Stacked Convolutional and Bidirectional Long Short-Term Memory. *IEEE Access* **2021**, *9*, 92600–92613. [[CrossRef](#)]
50. Haberbusch, M.; Bernardo, L.; Galassi, L.; Oddo, C.; Moscato, F. Electrocardiogram delineation using deep neural networks. *Stud. Health Technol. Inform.* **2022**, *293*, 117–118. [[CrossRef](#)]
51. Habib, A.; Karmakar, C.; Yearwood, J. Impact of ECG dataset diversity on generalization of CNN model for detecting QRS complex. *IEEE Access* **2019**, *7*, 93275–93285. [[CrossRef](#)]
52. Wu, W.; Huang, Y.; Wu, X. A new deep learning method with self-supervised learning for delineation of the electrocardiogram. *Entropy* **2022**, *24*, 1828. [[CrossRef](#)] [[PubMed](#)]
53. Chiang, H.T.; Hsieh, Y.Y.; Fu, S.W.; Hung, K.H.; Tsao, Y.; Chien, S.Y. Noise Reduction in ECG Signals Using Fully Convolutional Denoising Autoencoders. *IEEE Access* **2019**, *7*, 60806–60813. [[CrossRef](#)]
54. Ivanov, K.; Jekova, I.; Krasteva, V. Convolutional Autoencoder for Filtering of Power-Line Interference with Variable Amplitude and Frequency: Study of 12-Lead PTB-XL ECG Database. *Lect. Notes Netw. Syst.* **2023**, *658*, 121–133. [[CrossRef](#)]
55. Reljin, N.; Lazaro, J.; Hossain, M.B.; Noh, Y.S.; Cho, C.H.; Chon, K.H. Using the Redundant Convolutional Encoder–Decoder to Denoise QRS Complexes in ECG Signals Recorded with an Armband Wearable Device. *Sensors* **2020**, *20*, 4611. [[CrossRef](#)] [[PubMed](#)]
56. Prakash, A.J.; Patro, K.K.; Saunak, S.; Sasmal, P.; Kumari, P.L.; Geetamma, T. A New Approach of Transparent and Explainable Artificial Intelligence Technique for Patient-Specific ECG Beat Classification. *IEEE Sens. Lett.* **2023**, *7*, 5501604. [[CrossRef](#)]
57. Mehri, M.; Calmon, G.; Odille, F.; Oster, J. A deep learning architecture using 3D vectorcardiogram to detect R-peaks in ECG with enhanced precision. *Sensors* **2023**, *23*, 2288. [[CrossRef](#)] [[PubMed](#)]
58. Moskalenko, V.; Zolotykh, N.; Osipov, G. Deep Learning for ECG Segmentation. In *Advances in Neural Computation, Machine Learning, and Cognitive Research III. NEUROINFORMATICS 2019. Studies in Computational Intelligence*; Kryzhanovskiy, B., Dunin-Barkowski, W., Redko, V., Tiumentsev, Y., Eds.; Springer: Cham, Switzerland, 2020; Volume 856, pp. 246–254. [[CrossRef](#)]
59. Jimenez-Perez, G.; Alcaine, A.; Camara, O. Delineation of the electrocardiogram with a mixed-quality-annotations dataset using convolutional neural networks. *Sci. Rep.* **2021**, *11*, 863. [[CrossRef](#)] [[PubMed](#)]
60. Jimenez-Perez, G.; Alcaine, A.; Camara, O. U-Net Architecture for the Automatic Detection and Delineation of the Electrocardiogram. *Comput. Cardiol.* **2019**, *46*, 1–4. [[CrossRef](#)]
61. Zahid, M.; Kiranyaz, S.; Ince, T.; Devcioglu, O.; Chowdhury, M.; Khandakar, A.; Tahir, A.; Gabbouj, M. Robust R-peak detection in low-quality holter ECGs using 1D convolutional neural network. *IEEE Trans. Biomed. Eng.* **2022**, *69*, 119–128. [[CrossRef](#)] [[PubMed](#)]
62. Duraj, K.; Piaseczna, N.; Kostka, P.; Tkacz, E. Semantic segmentation of 12-lead ECG using 1D residual U-Net with squeeze-excitation blocks. *Appl. Sci.* **2022**, *12*, 3332. [[CrossRef](#)]
63. Liang, X.; Li, L.; Liu, Y.; Chen, D.; Wang, X.; Hu, S.; Wang, J.; Zhang, H.; Sun, C.; Liu, C. ECG\_SegNet: An ECG delineation model based on the encoder-decoder structure. *Comput. Biol. Med.* **2022**, *145*, 105445. [[CrossRef](#)] [[PubMed](#)]
64. Wang, D.; Qiu, L.; Zhu, W.; Dong, Y.; Zhang, H.; Chen, Y.; Wang, I. Inter-patient ECG characteristic wave detection based on convolutional neural network combined with transformer. *Biomed. Signal Process. Control* **2023**, *81*, 104436. [[CrossRef](#)]
65. Jia, D.; Zhao, W.; Li, Z.; Hu, J.; Yan, C.; Wang, H.; You, T. An electrocardiogram delineator via deep segmentation network. In Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 1913–1916. [[CrossRef](#)]

66. Jimenez-Perez, G.; Acosta, J.; Alcaine, A.; Camara, O. Generalizing electrocardiogram delineation: Training convolutional neural networks with synthetic data augmentation. *arXiv* **2021**, arXiv:2111.12996v1. [[CrossRef](#)]
67. Veit, A.; Wilber, M.; Belongie, S. Residual networks behave like ensembles of relatively shallow networks. In *Advances in Neural Information Processing Systems*; Neural Information Processing Systems: La Jolla, CA, USA, 2016; pp. 550–558.
68. van de Leur, R.; Bos, M.; Taha, K.; Sammani, A.; Yeung, M.; van Duijvenboden, S.; Lambiase, P.; Hassink, R.; van der Harst, P.; Doevendans, P.; et al. Improving explainability of deep neural network-based electrocardiogram interpretation using variational auto-encoders. *Eur. Heart J.-Digit. Health* **2022**, *3*, 390–404. [[CrossRef](#)] [[PubMed](#)]
69. Zyout, A.; Alquran, H.; Mustafa, W.A.; Alqudah, A.M. Advanced Time-Frequency Methods for ECG Waves Recognition. *Diagnostics* **2023**, *13*, 308. [[CrossRef](#)] [[PubMed](#)]
70. Bock, C.; Mortl, C.; Mahringer, C.; Huemer, M.; Meier, J. Variability of expert assessments of ECG time domain parameters. *Eur. J. Anaesthesiol. Intensive Care Med.* **2023**, *2*, e0020. [[CrossRef](#)]
71. Al-Karadi, M.S.; Langley, P. Multi-beat averaging reveals U waves are ubiquitous and standing tall at elevated heart rates following exercise. *Sensors* **2020**, *20*, 4029. [[CrossRef](#)] [[PubMed](#)]
72. Krasteva, V.; Jekova, I.; Leber, R.; Schmid, R.; Abächerli, R. Superiority of Classification Tree versus Cluster, Fuzzy and Discriminant Models in a Heartbeat Classification System. *PLoS ONE* **2015**, *10*, e0140123. [[CrossRef](#)] [[PubMed](#)]
73. Correa, R.O.; Laciari, E.; Valentinuzzi, M.E. Amplitude differences least squares method applied to temporal cardiac beat alignment. *J. Phys. Conf. Ser.* **2007**, *90*, 012073. [[CrossRef](#)]
74. Schiller, E.T.M. The Innovative ECG Analysis Program for the Clinical Application and Quality of ECG Analysis. Available online: <https://www.schiller.ch/en/software-connectivity/etm-s395> (accessed on 10 June 2024).
75. Kligfield, P.; Badilini, F.; Denjoy, I.; Babaeizadeh, S.; Clark, E.; De Bie, J. Comparison of automated interval measurements by widely used algorithms in digital electrocardiographs. *Am. Heart J.* **2018**, *200*, 1–10. [[CrossRef](#)]
76. Wagner, P.; Strodthoff, N.; Boussejot, R.; Samek, W.; Schaeffter, T. PTB-XL, a large publicly available electrocardiography dataset (version 1.0.1). *PhysioNet* **2020**. [[CrossRef](#)]
77. Wagner, P.; Strodthoff, N.; Boussejot, R.D.; Kreisler, D.; Lunze, F.; Samek, W.; Schaeffter, T. PTB-XL, a large publicly available electrocardiography dataset. *Sci. Data* **2020**, *7*, 154. [[CrossRef](#)] [[PubMed](#)]
78. Moody, G.B.; Koch, H.; Steinhoff, U. The PhysioNet/Computers in Cardiology Challenge 2006: QT interval measurement. In *Proceedings of the 2006 Computers in Cardiology, Valencia, Spain, 17–20 September 2006*; pp. 313–316.
79. Boussejot, R.; Kreisler, D.; Schnabel, A. Nutzung der EKG-signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik* **1995**, *40*, S317. [[CrossRef](#)]
80. Goldberger, A.; Amaral, L.; Glass, L.; Hausdorff, J.; Ivanov, P.; Mark, R.; Stanley, H. PhysioBank PhysioToolkit and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, e215–e220. [[CrossRef](#)] [[PubMed](#)]
81. Perez Alday, E.A.; Gu, A.; Shah, A.; Liu, C.; Sharma, A.; Seyedi, S.; Bahrami Rad, A.; Reyna, M.; Clifford, G. Classification of 12-lead ECGs: The PhysioNet/Computing in Cardiology Challenge 2020 (version 1.0.2, 2022). *Physiol. Meas.* **2020**, *41*, 124003. [[CrossRef](#)] [[PubMed](#)]
82. 68–95–99.7 Rule. Wikipedia. Available online: [https://en.wikipedia.org/wiki/68-95-99.7\\_rule](https://en.wikipedia.org/wiki/68-95-99.7_rule) (accessed on 1 April 2024).
83. Willems, J.L.; Arnaud, P.; van Bommel, J.H.; Degani, R.; Macfarlane, P.W.; Zywiec, C. Common standards for quantitative electrocardiography: Goals and main results. *Methods Inf. Med.* **1990**, *29*, 263–271. [[PubMed](#)]
84. The CSE Working Party. Common Standards for Quantitative Electrocardiography. CD-ROM Version of the CSE Database. 1990. Available online: <https://www.medteq.net/article/2017/4/1/iec-60601-2-25-cse-database-test-experience> (accessed on 1 April 2024).
85. IEC60601-2-51:2003; Medical Electrical Equipment—Part 2–51: Particular Requirements for Safety, including Essential Performance, of Recording and Analysing Single Channel and Multichannel Electrocardiographs. International Electrotechnical Commission: London, UK, 2003.
86. Turchenko, V.; Chalmers, E.; Luczak, A. A deep convolutional auto-encoder with pooling—Unpooling layers in Caffe. *Int. J. Comput.* **2019**, *18*, 8–31. Available online: <https://www.computingonline.net/computing/article/view/1270/840> (accessed on 1 April 2024). [[CrossRef](#)]
87. Siddique, N.; Paheding, S.; Elkin, C.; Devabhaktuni, V. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* **2021**, *9*, 82031–82057. [[CrossRef](#)]
88. Götz, T.I.; Göb, S.; Sawant, S.; Erick, X.F.; Wittenberg, T.; Schmidkonz, C.; Tomé, A.M.; Lang, E.W. A Ramming Number of necessary training examples for Neural Networks with different number of trainable parameters. *J. Pathol. Inform.* **2022**, *13*, 100114. [[CrossRef](#)]
89. Willems, J.L. *The CSE Multilead Atlas Manual-Measurement Results Dataset-3*; ACCO: Leuven, Belgium, 1988.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.