A Semantic Segmentation-based Digitization of ECG Papers

Davyd Melo[1], João Paulo Madeiro[1], Luís O. Rigo Jr.[2], Cláudia do Ó Pessoa[1], José Antônio Macêdo[1], Danielo G. Gomes[1]

[1] Federal University of Ceará, Fortaleza, Brazil
[2] University for International Integration of the Afro-Brazilian Lusophony, Redenção, Brazil

Abstract

The electrocardiogram (ECG) is crucial to identify biomarkers of cardiovascular diseases (CVD) in a non-invasive and non-expensive manner. While digital ECGs have become more prevalent and offer many advantages, the use of ECGs papers may still be necessary in certain situations, e.g. equipment failure, accessibility, familiarity, legal requirements. Healthcare providers must weigh the benefits and limitations of each method and choose the most appropriate option for their patients. Once ECG papers are converted into digital format, they can be used to teach machine learning models how to accurately diagnose heart conditions. By using a collection of digital ECGs alongside their corresponding diagnoses, healthcare providers can help train these models to automate the diagnostic process, making it faster and more efficient. This paper proposes a digitization pipeline that uses a combination of machine learning and image processing techniques. The system identifies the regions of interest on the scanned papers and segments the ECG signal from the background, by using a Random Forest model trained using a Gabor filter bank. Finally, the ECG waveform in millivolt units is extracted through an iterative process that uses the ECG baseline as a reference. After testing the proposed digitization pipeline on two sets of images, the results were accurate (MSE = 0.021). Also, the estimations were closely matched the actual ECG data (Correlation Coefficient = 0.861).

1. Introduction

The electrocardiogram (ECG) is a simple, low cost and non-invasive test that evaluates the electrical activity of the heart. During the systolic and diastolic processes, a series of electrodes capture the potential difference between some points on the chest and a reference on the limbs. The waveforms provided by the ECG test have been extensively used for the diagnosis of some CVDs, thus helping doctors to correctly assist the patients and, consequently, save lives [1].

Although certain electrocardiogram (ECG) equipment does generate digital output, it is a prevailing practice within clinical and hospital settings to utilize cost-effective devices that exclusively produce printed ECG reports. Consequently, this approach results in challenges related to the storage, transmission, and processing of physical paper-based ECG records.

An additional issue stemming from the exclusive reliance on physical ECG printouts is the associated opportunity cost of neglecting data processing. Employing a digital ECG format affords the potential for substantial reductions in diagnosis time through the application of signal processing tools, which can assist healthcare practitioners in their clinical workflows. Moreover, this approach offers the benefit of constructing datasets suitable for integration into machine learning algorithms capable of extrapolating insights from historical data, thereby enhancing diagnostic capabilities.

Hence, this study presents a semantic segmentation-based technique designed for the extraction of digital signals from scanned ECG paper records. This methodology serves to mitigate the prevalence of free parameters inherent in image processing algorithms, which typically necessitate manual adjustments for optimal functionality.

2. Materials

2.1. MIT-BIH Arrhythmia Dataset

MIT-BIH arrhythmia database is a widely used database for building and testing arrhythmia detectors [2]. It contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects. The dataset's webpage provides a web application called Lightwave that displays the ECG data samples as SVG images. For each subject in the dataset, a scraping algorithm is executed to extract an image corresponding to a ten-second excerpt. A subset of 10 images were selected for the experiments. These extracted images are converted to PDF format for scanning at 300 dpi resolution. The data files, also present in the database's webpage, are used as ground truth for performance evaluation of the digitizing process.

2.2. Samsung Health Dataset

Recently, wearable devices, like smartwatches, are fully equipped with sensors capable of reading biological data from their users [3]. Samsung Watch5 provides sensors like ECG, PPG, and so on. The Samsung Health App can export a 30-second ECG signal using PDF and CSV formats. A set of 21 ECG papers are collected from one of the authors and two inpatient of an associated hospital. The study protocol was approved by the Human Research Ethics Committee of the Universitary Hospital Walter Cantídio (CEP/HUWC) – CAAE: 69745323.3.0000.5045. The PDF files are printed and then scanned with a 300 dpi resolution while the CSV files are used as ground truth. For simplicity and to avoid noisy data on the beginning of the data, only the samples from 10 to 20 seconds of each image are used.

3. Methodology

In order to acquire the ECG signal in a 1d-array format, the images step through the digitizing pipeline shown in Figure 1. The next subsections delineate each block on the proposed architecture.

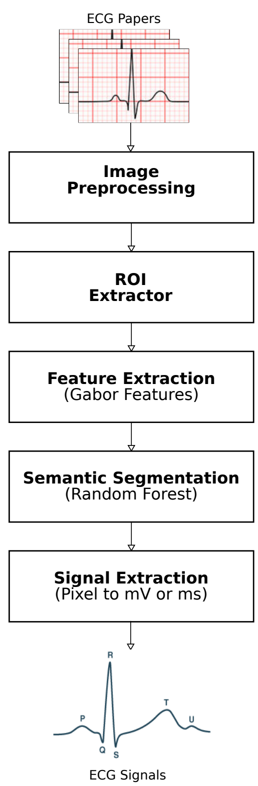

Figure 1. Proposed digitization pipeline

3.1. Pre-processing

Initially, to remove the scanning marks on the borders, the images are trimmed on the edges by 50 pixels factor. After that, a 5x5 normalized box filter applies blurring to remove the high frequency components from images, reducing the grayscale level variability. This last step improves the line detection for baseline identification and helps the pixel labeling process.

3.2. Region of interest detection

Each scanned image can contain multiple ECG leads that can be separated and processed individually. The grayscale levels of the background have a significant difference from the grayscale level of the ECG signals. Thus, to determine the bounding box of each lead, a sliding windowing is applied from each side of the paper and calculates the variance of the grayscale levels. For each step of the window, the total variance is obtained by summing all window elements. When this accumulated variance reaches a maximum, a bounding is detected.

3.3. Feature extraction

The feature extraction phase is based on a bank of Gabor filters. These filters are extensively used on texture segmentation tasks [4] [5]. A Gabor filter is a band-pass filter composed by a sinusial function multiplied by a gaussian envelope. In complex domain, the 2D Gabor filter is defined like:

$$
\begin{aligned}
g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \\
exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right)
\end{aligned}
\tag{1}
$$

where $x' = x\cos(\theta) + y\sin(\theta)$ and $y' = -x\sin(\theta) + y\cos(\theta)$. In the first part of 1, σ and γ controls the amount of spread and the ellipticity of the gaussian envelope, respectively. Lastly, λ controls the wavelength of the oscillatory function and ψ is a phase factor.

To construct the filter bank, the parameters of the Gabor filter are varied and combined with each other inside of the following ranges: $\lambda \in \{0, \pi/4, 2\pi/4, 3\pi/4\}$, $\theta \in \{0, \pi/4\}$, $\sigma \in \{1, 3\}$ and $\gamma \in \{0.05, 0.5\}$. This way, set of kernels are created with different frequencies and orientations, allowing the analysis of the images at multiple levels of detail.

After the construction of the filter bank, each kernel is convolved with the image producing a new filtered image. This filtered image is converted in a vector through a vertical stacking process. Theses column vectors, extracted by each kernel, are stacked horizontally to create a feature matrix that compose the actual dataset used later for training and test purposes.

3.4. Semantic segmentation

In computing vision research, Semantic Segmentation is a process that assigns a label to each pixel in an image, making it more understandable for posterior tasks [6]. To extract the ECG, a Random Forest (RF) model separates, through semantic segmentation, the signal, labeled with 1, from other paper elements, like grid lines or textual elements, labeled with 0. The training images are manually labeled, with the aid of free image editing software, identifying the range of grayscale levels of the ECG signal.

The Random Forest (RF) is an ensemble machine learning method that builds upon a set of Decision Trees (DT) to reduce prediction bias and overfitting [7]. RF relies on bagging to select a different subset of data for training each DT but also improves the randomness by selecting the best subset of predictors, what is called random subspace method or feature bagging [8]. For classification problems, like semantic segmentation, the underlying predictions of each DT are aggregated using a majority vote approach. The selected RF model trains ten decision trees that make data splits using the Gini impurity criterion. Feature bagging considers the square root of the total amount of features for the maximum size of feature splits.

3.5. Signal extraction algorithm

The extraction of the ECG signal from the segmented image follows a strategy based on the signal's baseline. The baseline constitutes the starting point of an iterative algorithm to select the most suitable pixels for each column of the segmented image. To extract the baseline, a Canny edge detector followed by a probabilistic Hough transform is applied to detect horizontal lines. It allowed a little slack of 2° in the slope of the lines. The baseline value is defined as the median value of the y coordinates of the midpoint of each line.

After the baseline discovery, an iterative algorithm searches for a white pixel in each column of the segmented image. At the first iteration, each column is traversed in both directions, starting from the pixel that corresponds to the baseline. The pixel height with a smaller distance from the baseline is stored and feeds the starting point for the next iteration. A total of 20 iterations were executed, producing an array containing the heights of the white pixels in the segmented mask. Finally, the pixel heights array is resampled in the x-axis (3600 and 5000 samples for the MIT-BIH and Samsung Health datasets, respectively) and scaled down based on ECG paper's block size in pixels, producing an array with floating-point values in mV units.

	Dataset	
Metric	**MIT-BIH**	**Samsung Health**
Mean Squared Error	0.035 ± 0.032	0.007 ± 0.005
Pearson Correlation	0.833 ± 0.122	0.888 ± 0.071
Heart Rate (MAE)	1.457 ± 3.617	0.115 ± 0.191
R-wave Peak (MAE)	2.973 ± 2.124	2.690 ± 1.896

Table 1. Performance metrics used to evaluated the digitization method.

4. Results

The proposed method for digitizing electrocardiograms (ECG) was subjected to validation procedures against two distinct test sets, which were not exposed to the Random Forest (RF) model during its training phase. A combined total of 10 and 20 ECG images were drawn from the MIT-BIH Arrhythmia Dataset and the Samsung Health Dataset, respectively. Various metrics were computed to comprehensively evaluate each phase of the processing pipeline.

For the assessment of the semantic segmentation phase, the DICE coefficient above 0.998 was obtained for both datasets. The performance evaluation of the signal extraction phase was based on two metrics, namely, the Mean Squared Error (MSE) and Pearson's correlation coefficient. These metrics were employed to quantitatively gauge the degree of similarity between the estimated signal and the ground truth signal.

Additionally, the Heart Rate (HR) and R-peaks positions were extracted to provide practical and clinically relevant metrics. To evaluate how close these metrics are on ground truth and extracted signals, the Mean Absolute Error (MAE) was computed. These metrics are of significant importance in the field of medical practice, as they are routinely employed by healthcare professionals. Table 1 summarizes the metrics procured over the course of the experimental study. Furthermore, Figure 2 depict the Bland–Altman plot of the sample positions of the R-peaks from the predicted and ground-truth signals (MIT-BIH dataset). The peaks were extracted using NeuroKit2 library [9]. Figure 3 shows an excerpt of an signal extracted using the digitization process versus the ground truth signal.

5. Discussion

From the metrics table it is possible to notice that the proposed pipeline achieves low MSE and a high Person correlation, maintaining a good alignment between the extracted signals and ground truth signals. The metrics of Heart Rate (MAE) and R-wave Peak (MAE) evince that the digitization procedure maintained the frequency information pertaining to heart rate, with an acceptable margin of deviation in the localization of R-peaks. Furthermore,

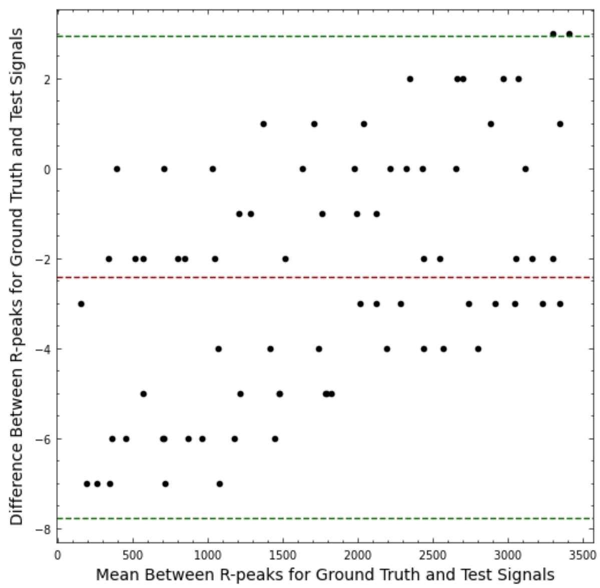

Figure 2. Bland-Altman plot for the R-peak locations (Samsung Health dataset).

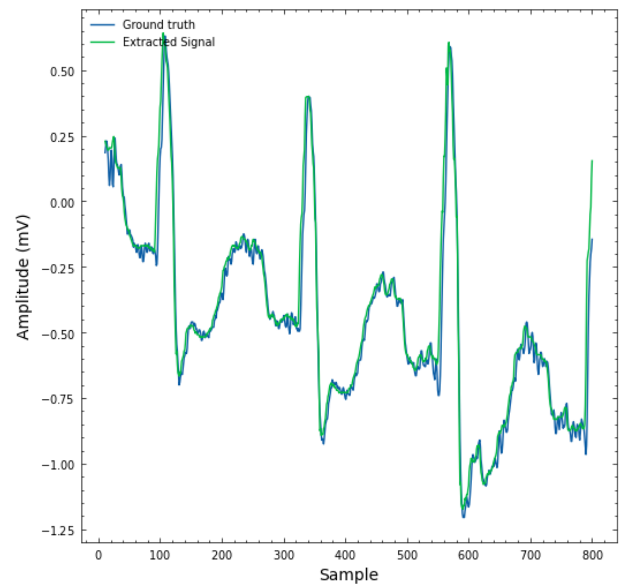

Figure 3. Extracted signal (green) ploted over the ground truth signal (blue).

the Bland-Altman plot reveals that the majority of data points are situated within the Limits of Agreement (LoA), in close proximity to the line delineating the mean differences. This observation signifies a concordance between the estimated values and the values deemed as actual.

6. Conclusions

ECG papers constitute a very rich source of information that can be used to train machine learning models that can aid physicians to gain insights about CVDs. This paper proposes a digitization pipeline that uses a combination of machine learning and image processing techniques. Results show that the proposed pipeline digitizes ECG papers with low error, high correlation and maintain the heart rate when compared with the ground truth signals.

Acknowledgments

Some of the results presented in this work were obtained through the project “CENTRO DE EXCELÊNCIA EM INTELIGÊNCIA ARTIFICIAL – AI4WELLNESS”, funded by Samsung Eletrônica da Amazônia Ltda., under the scope of law No. 8.248/91 (Lei de Informática).

References

[1] Surawicz B, Knilans T. Chou’s electrocardiography in clinical practice: adult and pediatric. Elsevier Health Sciences, 2008.

[2] Moody GB, Mark RG. The impact of the mit-bih arrhythmia database. IEEE engineering in medicine and biology magazine 2001;20(3):45–50.

[3] Iqbal SM, Mahgoub I, Du E, Leavitt MA, Asghar W. Advances in healthcare wearable devices. NPJ Flexible Electronics 2021;5(1):9.

[4] Idrissa M, Acheroy M. Texture classification using gabor filters. Pattern Recognition Letters 2002;23(9):1095–1102.

[5] Armi L, Fekri-Ershad S. Texture image analysis and texture classification methods-a review. arXiv preprint arXiv190406554 2019;.

[6] Lateef F, Ruichek Y. Survey on semantic segmentation using deep learning techniques. Neurocomputing 2019;338:321–348.

[7] Breiman L. Random forests. Machine learning 2001;45:5–32.

[8] Ho TK. The random subspace method for constructing decision forests. IEEE transactions on pattern analysis and machine intelligence 1998;20(8):832–844.

[9] Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SHA. NeuroKit2: A python toolbox for neurophysiological signal processing. Behavior Research Methods feb 2021;53(4):1689–1696. URL

Address for correspondence:

Davyd Bandeira de Melo
Bibio Frota St, 505, apt. 503A, Fortaleza-CE, Brazil
davydmelo@gmail.com