Towards assisted electrocardiogram interpretation using an 
AI-enabled Augmented Reality Headset

P. Lampreave\textsuperscript{a}, G. Jimenez-Perez\textsuperscript{a}, I. Sanz\textsuperscript{b}, A. Gomez\textsuperscript{c} and Oscar Camara\textsuperscript{a}

\textsuperscript{a}BCN MedTech, PhySense Group, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain; \textsuperscript{b}Servicio de Medicina Interna, Hospital Vall d’Hebron, Barcelona, Spain; \textsuperscript{c}King’s College London, London SE1 7EH, UK. School of Biomedical Engineering & Imaging Sciences

\textbf{ARTICLE HISTORY}
Compiled July 13, 2020

\textbf{ABSTRACT}
The interpretation of surface electrocardiograms (ECGs) is key for the diagnosis and monitoring of cardiovascular health and optimal clinical decision making. Despite the progressive digital transformation in healthcare, it is still common for clinicians to visualize and analyze ECG printed on paper. Although some acquisition systems provide a signal processing-based classification of the ECG, clinicians often find it unreliable and tend to ignore it. During the last few years, Artificial Intelligence (AI) techniques have become state-of-the-art for ECG processing. The organization of analysis challenges using curated ECG databases (i.e. PhysioNet Resource), has highly contributed to the proliferation of AI-based ECG analysis tools. Nevertheless, the lack of digitised ECG in real clinical environments has hampered the clinical translation of these techniques. Concurrently, we are living a rise in virtual and augmented reality (AR) technologies, with an increasing availability of devices that can also have an impact in clinical workflows. In this work, we present a computational framework for the automatic digitisation and assisted interpretation of ECG based on an AI-enabled Augmented Reality (AR) headset. The AR headset is used to acquire an image of the printed ECG, which is then processed to extract the digitised ECG signal. Afterwards, the extracted ECG is introduced into a Deep Learning (DL) algorithm pre-trained on the PhysioNet/CinC 2020 Challenge database, composed of 12-lead ECG recordings of 6877 patients with nine cardiomyopathy categories. The output of the DL algorithm classifies the ECG signal onto the different categories, which is then visualized back in the AR headset. Preliminary classification results on simulated ECG images (96.5% of accuracy, obtained nearly in real-time) and initial feedback from physicians in emergency departments confirm the potential of the developed approach to contribute on the digital transformation of ECG processing.

\textbf{KEYWORDS}
Electrocardiogram; medical data digitisation; augmented reality; deep learning

1. Introduction

An electrocardiogram (ECG) is a clinical test used to measure the heart’s rhythm and electrical activity. Information provided by an ECG is used in the diagnosis and monitoring of different heart illnesses such as blocked arteries or pathological arrhyth-
mias. The test involves the use of electrodes, cables that conduct the small signals recorded and an electrocardiogram machine that is able to amplify the signal, filter out any undesired noise and display (on a monitor) or print on paper (López Farré and Macaya Miguel 2009) the recorded ECG waves. A major issue in ECG interpretation is the high inter-patient and inter-operator variability of the captured signal due to the placement of the electrodes, the morphology of the heart, the type of pathology and the patient (Rajaganeshan et al. 2008). This issue is partly circumvented if a highly experienced physician carries out the interpretation; however, this is not always possible and some ECG features might go unnoticed to non-specialists or even when an expert cardiologist is available, especially when analysing multiple leads for several heart cycles or in stress-related situations. Although most ECG acquisition systems include basic signal processing algorithms that provide an initial ECG interpretation, these methods analyse ECG signals as they are acquired, lacking global vision of the entire trace and producing generally unreliable results, hence clinicians tend to ignore them. As a result, high-quality assisted ECG interpretation is still an unmet need and may be crucial to enable reliable and fast readings for correct clinical decisions.

Despite the increasing digitisation of medical records and diagnostic tests in health-care institutions, there is still a substantial number of procedures relying on analog data, as it is the case of ECG analysis; clinicians usually diagnose a patient performing a visual inspection of a paper-printed ECG, which is later stored in a paper archive. If a digital version of the ECG is available, modern computational methods can help unburden clinicians of complex pattern detection tasks by providing objective measurements over clinical data (Mincholé et al. 2019) or by aiding in the discovery of potential biomarkers (Lyon et al. 2018; Faust et al. 2018). Historically, digital signal processing algorithms have been successfully applied to ECG analysis such as using wavelet transforms for ECG delineation (Pablo Martínez et al. 2004). However, these methods require laborious rule adaptation when extended to cardiac morphologies not represented in the training dataset. More recently, Artificial Intelligence (AI) and Machine Learning techniques are being adapted to ECG processing, mainly for wave classification (Hou et al. 2019; Hannun et al. 2019), but also for ECG delineation (Jimenez-Perez et al. 2020). The availability of annotated databases of ECG recordings corresponding to different cardiac patients such as the PhysioNet Resource (MIT Laboratory for Computational Physiology) (Goldberger et al. 2000) has significantly contributed to the development of AI-based ECG processing algorithms.

Virtual and Augmented Reality (AR) have enabled more intuitive and effective ways of visualizing and interacting with complex medical data, with potential in education, training, pre-operative planning and intra-operative support. In particular, AR devices are interesting in healthcare since they can provide extra information without disengaging the clinical user from the physical context, which is essential in most clinical situations e.g. operating or emergency rooms. For instance, Okamoto et al. (2015) projected pre-operative medical images during the actual procedure to better locate organs of interest. In the cardiology field, Lamounier et al. (2010) recorded an ECG and projected a beating heart on a patient’s chest using markers, for educational purposes. More recently, Jang et al. (2018) developed a framework for 3D holographic visualization of myocardial scar, imaged using magnetic resonance (MR) imaging, on the HoloLens AR device, to facilitate MR-guided ventricular tachycardia ablation. AR devices can also be used to visualize and digitise medical data, analysing the scene capture by the headset, as demonstrated in (Treivase et al. 2019), where they proposed a generic framework to extract ultrasound (US) images and superimpose the results of an analysis task (i.e. based on state-of-the-art techniques such as Deep Learning),
without any need for physical connection or alteration to the US system.

In this paper, we propose a novel computational framework towards automatic digitisation and assisted interpretation of paper-based ECGs using an AI-enabled Augmented Reality (AR) headset. The developed framework was tested on simulated ECG datasets built from the PhysioNet/CinC 2020 challenge database (Perez Alday et al. 2020) of more than six thousand ECG recordings.

2. Methods

An overview of the proposed methodology is illustrated in Figure 1, with each step of the process being described in more detail below. First, the built-in camera of the AR headset is used to acquire an image of the printed ECG (Section 2.1). Subsequently, the acquired image is processed to extract the 1D ECG signal (Section 2.2). This signal is then fed to a deep learning (DL) algorithm previously trained with an ECG database available in PhysioNet, that classifies the ECG signal onto 9 different categories (Section 2.3). Last, the classification result is visualized back in the AR headset (Section 2.4).

![Figure 1. Scheme of the proposed methodology. An image of the paper ECG is captured with the camera of the AR headset. This image is processed to extract the digital ECG signal, which is then classified using deep learning into different cardiomyopathy categories. Finally, the classification result is displayed on the AR headset together with the real paper ECG.](image-url)
2.1. ECG capture with AR headset

The Meta 2 AR headset (see Section 3) was programmed to acquire an image with its built-in camera at the press of a mounted button using Unity 3D. A Unity Scene was created using the MetaCameraRig object for interacting with the headset, the ButtonEventBroadcaster callback for modifying the behaviour of the rightmost headset button and the MetaButtonWebCamStillImageController object to save the headset’s field of view as a PNG image and to trigger the rest of the processing pipeline. Figure 2 shows an example of captured ECG image.

![Example of an image of the printed ECG recording acquired with the Augmented Reality headset. Four green dots in the corners were manually marked to facilitate the automatic ECG extraction.](image)

2.2. ECG signal extraction

The acquired image was post-processed to segment the 1D ECG recording using the OpenCV library (Bradski 2000). For this purpose, a multi-step pre-processing pipeline was applied to retrieve the signal, consisting in a region of interest (ROI) boundary retrieval step for performing a cropping operation, a grid removal step, and a final sampling value computation step. The pipeline’s steps are represented in Figure 3.

To facilitate reliable extraction of the ECG, we indicated the extent of the ECG with four green marks on the paper (Figure 2). These markers were localized in the image as follows: firstly, regions of high saturation were selected by thresholding the image in the HSV color scheme. Secondly, noisy activations were removed through the usage of a median blur filter. Thirdly, a dilation operation was performed to fill any holes in the binary mask. Finally, the coordinates of each fiducial were computed through the computation of the centroids each connected component. These coordinates were used to crop the image, isolating the ROI.

A second step consisted in the removal of the background grid. For this purpose, the black colored pixels (text and waves) were threshold-selected and filtered out. Then, the image was converted to the L*a*b color scheme to achieve near-uniform spacing of perceived color differences. The a and b channels (corresponding to red/green and blue/yellow, respectively) were inputted into a k-means algorithm for separating the input data into $k = 3$ clusters, corresponding to the background, the red grid and the black signal. All pixels corresponding to the signal cluster were isolated and cropped.

---

1 https://unity3d.com
Figure 3. Scheme of the processing pipeline to extract the ECG signal from the image acquired with the Augmented Reality headset.

into independent leads. Given similar positions for the label names across all signals, these were deleted.

In the third and final step, the value of the signal at a specific pixel was assigned to be that of the median row location for each column in the lead-cropped image. Missing values were substituted by cubic splines interpolation of neighbouring samples (Meyer and Keiser 1977; Jane et al. 1992). Figure 4 shows an example of the different processing steps to extract the ECG signals from an image containing three leads, including intermediate results.

2.3. Neural network architecture

The digitised ECG signal was fed into a convolutional neural network (CNN) for classification. The network architecture comprised a convolutional feature extractor and a fully connected discriminator for producing the prediction (LeCun et al. 2015), as it is illustrated in Figure 5. The convolutional feature extractor was composed of eight convolutional layers ($ks = 3$, zero-padding), each but the first followed by a max pooling operation (after which the number of filters was doubled, starting at 16) for promoting translation invariance and leaky ReLU non-linearities. The convolutional extractor used batch normalization (Ioffe and Szegedy 2015) and spatial dropout (Tompson et al. 2015) with probability $p = 0.1$ as regularizer after the non-linearity. The extracted features were summarized using a global average pooling (GAP) operation and introduced into the classification block, which was composed of 5 trainable linear layers: an input of 256 features (GAP operation), which were mapped to 128, 64, 32 and 9 features. All operations were followed by leaky ReLU activations, batch normalization and dropout ($p = 0.1$).

The network was trained using the multi-label soft margin loss function and optimized with stochastic gradient descent (Bottou 2010), with a starting learning rate of 0.001 and Nesterov momentum to escape local minima, over 100 epochs, halving the learning rate on plateau if the performance did not improve after 10 epochs, and early stopping was applied to retrieve the best performing model on the validation set after training. The network was implemented using PyTorch (Paszke et al. 2019).
Figure 4. Example of different processing steps to extract the ECG signals from the image containing three leads. a) Original image; b) Conversion to HSV for marker detection and cropping; c) ECG signal extraction after color thresholding and k-means clustering; d) Zoom of the extracted ECG signals superimposed with original image; e) Zoom of the extracted ECG signals after cubic interpolation superimposed with original image.

2.4. Classification visualization in AR headset

The resulting ECG classification (classification probabilities associated to each class) was visualized with the AR headset (see Figure 6). A new Unity Scene was created using the MetaCameraRig to register the display to the field of view of the AR headset. The actual GUI was designed as a big canvas, which was used as a base for the rest of
the GameObjects that conformed the scene: for each class, a text box that gives the actual percentage of the prediction, a slider providing a more visual representation of the results, and a text box with the name of the diagnosis. This environment was loaded after the image acquisition and set to sleep until the classifier returned the probabilities returned by the classifier.

3. Materials and Data

We used the Meta2 (Metavision) AR headset, which has two programmable physical buttons to allow user interaction. The AR interface was implemented in Unity3D using a Unity SDK provided by Metavision.

ECG data from the PhysioNet/CinC 2020 Challenge (Perez Alday et al. 2020) was used, specifically the first released data batch corresponding to the China Physiological Signal Challenge 2018 (Liu et al. 2018). The database consists of 12-lead ECG recordings of 6877 patients (46.21% female) collected from 11 hospitals. The recordings were sampled at 500 Hz, lasting between six and sixty seconds, and were classified into 9 possible cardiomyopathies, including Atrial fibrillation (AF), First-degree atrioventricular block (I-AVB), Left bundle branch block (LBBB), Right bundle branch block (RBBB), Premature atrial contraction (PAC), Premature ventricular contraction (PVC), ST-segment depression (STD) and ST-segment elevation (STE).

3.1. Data preparation for ECG extraction

The current circumstances (i.e. the COVID-19 crisis of 2020) prevented us from curating our existing paper-based ECG database. As a result, we simulated paper-like ECG traces from the PhysioNet database, as follows. Using the ecg-plot Python package, we produced ECG traces akin to those available from our clinical collaborators, after setting the color of the ECG wave to black, and using only three leads per chart. Four green dots at the corners were also added as described in Section 2.2. The resulting ECG was displayed in a monitor (see Figure 6), resembling our real paper-based ECG traces (as shown in Figure 2).

6860 samples from the selected database were finally used for this purpose, as 17 recordings produced errors during the signal extraction phase, preventing their usage.

---

2https://pypi.org/project/ecg-plot
in the classification model. The simulated recordings corresponded to the samples used for training and validating the machine learning model.

### 3.2. Data preparation for ECG classification

The dataset was split in a subject-wise manner (Faust et al. 2018), grouping all recordings of the same patient as well any further subdivisions (windows, beats) in the same set. Given that no test set was needed because of the nature of the challenge, the dataset was split uniquely into a training and validation sets (75%-25%) in a stratified manner. The recordings were pre-processed in order to train the network: a low-pass filter at \( f = 125 \text{Hz} \) was applied to reduce noise and a high-pass filter at \( f = 0.5 \text{Hz} \) was applied to suppress baseline wander. Furthermore, the recordings were reshaped to contain 18432 samples, zero-padding or cropping when necessary.

For increasing the generalization of the trained network, data augmentation was applied to the initial training set for reducing the model’s variance. This work features ECG-tailored data augmentation, partially based on (Jimenez-Perez et al. 2020), which was amplitude-related to the input signal through a signal-to-noise ratio; specifically, additive white Gaussian noise (SNR = 20 dB), powerline and baseline noise (20 dB at 50 Hz and 5 dB at 0.5 Hz, respectively) and random amplitude changes (1% of the original signal amplitude) were considered. To avoid strict hyperparameters, every call to the data transformation code was accompanied by slight modifications in the SNR through a draw from an uniform distribution in the range of \( \pm 10\% \) of the original SNR.

### 4. Results

The ECG signal extraction pipeline was successful in the majority of the 6877 processed traces (such as the example displayed in the figure, where most of the ECG traces are recovered, with minor amplitude differences in some ECG peaks), with only 17 major failures related to incorrect image cropping or bad interpolations due to an insufficient number of samples recovered from the image. The simulated recordings were generated at a fifth of the original sampling frequency, coherently with the low resolution of printed ECGs, and were subsequently upsampled to the original 500 Hz.

The 12-Lead ECG was then used as an input for the DL-based classification model described in Section 2.3, outputting a prediction of the likelihood of belonging to a one of the nine target cardiomyopathies. To evaluate the performance of the pipeline on the pre-trained classifier, we employed the metrics proposed by the PhysioNet/CinC 2020 challenge organisers: the accuracy, which reflects the number of correct predictions in the whole dataset; the F-measure, computed as the harmonic mean of precision and recall; and the Fbeta-measure, with a beta value equal to two (i.e. F2-measure), which gives more weight to recall rather than precision.

One way to indirectly assess the impact of the ECG signal simulation and extraction procedures is to compare the classification metrics when having as input the original digital signals provided to the challengers versus the signals extracted with the proposed pipeline. The obtained results are summarized in the Table 1, where the classification metrics are only slightly lower for the extracted signals compared to the original ones. Table 1 summarizes the classifier’s performance when comparing these two approaches.

Once the classification of a given ECG recording was available, the obtained results
Table 1. Accuracy metrics for the classification of ECG signals.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>F-measure</th>
<th>F2-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital signals</td>
<td>0.983</td>
<td>0.906</td>
<td>0.892</td>
</tr>
<tr>
<td>Extracted signals</td>
<td>0.965</td>
<td>0.806</td>
<td>0.791</td>
</tr>
</tbody>
</table>

were displayed in the AR headset, as illustrated in Figure 6. In the example shown in this figure, it can be appreciated the simultaneous visualization of the ECG acquired with the AR system together with the classification results, in this case suggesting Atrial Fibrillation as the most likely cardiomyopathy. The visualization of the classification results are nearly real-time, with a minor delay coming from the calibration process of the AR headset; signal extraction and classification tasks take less than a second to run, once the network is trained.

5. Discussion and conclusions

Machine-aided diagnostic pipelines are growing in popularity given the continuous technological advancements and increasing digitisation in hospital and primary care centers. However, small databases and lack of expert knowledge and time to annotate them are constantly hindering the clinical translation of these state-of-the-art tools. In electrocardiography these advancements are further hampered by lack of digital support (Sassi et al. 2017) and a strong need for interpretable solutions (Faes et al. 2019) that effectively aid physicians during routine practice or in stress-related scenarios such as in emergency departments. Although AI-based pipelines have reduced performance gaps between automatic diagnostic tools and physicians by a wide margin, there are very few technological solutions in the literature that move towards a ‘man-plus-machine’ approach (Faes et al. 2019), effectively bridging the gap that prevents the widespread adoption of these technologies.

Figure 6. Example of ECG classification result (in this case Atrial Fibrillation (AF) with a probability of 71%, displayed in the Augmented Reality headset together with the visualization of the real ECG recording.
This clear need for seamless integration of diagnostic aid technologies is an opportunity for the rise and adoption of augmented reality technologies. The ability to interact with the physical medium, capture clinical information and visualize outputs of state-of-the-art analysis tools can revolutionize the clinical environment, drastically reducing the physician’s workload. In this work, we have presented a methodological modular-based pipeline for scanning printed ECGs “on-the-fly” for their digitisation and posterior classification into the most likely cardiomyopathy. This pipeline effectively recovers recordings from the digital device for accessing a plethora of downstream applications, in this case through a deep learning model for assisted diagnosis. The execution of the pipeline is in nearly real-time, which allows for its application to any downstream task with reasonable performance. The developed pipeline’s performance produced deviations of about 10% in $F_1$ score with respect to original digital data in the DL-based ECG classification. Although the performance gap is not negligible, the obtained results demonstrate the feasibility of the proposed pipeline, with potential benefits outweighing the drawbacks.

The current pipeline implementation, however, has certain limitations that hinder its performance and hamper its direct applicability. On the first hand, the dependence of the algorithm on the correct location of the fiducials (green markers) as well as in the correct acquisition and sufficient quality of the input image (e.g. issues with paper rotation, insufficient illumination, partial occlusions, image blurring due to degradation of the physical format, among others). Some of these issues could be solved from an algorithmic perspective, e.g. through the application of fiducial placement/strip segmentation algorithms, but others are entirely dependent on the quality of the acquired images. On the other hand, the use of more up-to-date AR headsets (e.g. Magic Leap, Hololens2) would facilitate the streamlining and connection between the different systems and allow more sophisticated services and applications.

On another note, the proposed solution presents independent signal extraction and classification tasks; an end-to-end system, trained directly with the ECG scanned images could potentially produce similar results. Because of the modularity of our work, this model could replace the extraction and classification steps without changing the acquisition and presentation results. Another direction for expansion is the usage of architectural changes, such as attention operations (Prabhakararaao and Dandapat 2020), or post-hoc interpretation tools, such as class activation maps (Yang et al. 2019), which allow the visualization of the features that have led the network to produce a specific diagnostic. Although current technologies are increasingly black-box given the advent of DL, the application of these post-hoc analysis tools can have a multiplicative effect in a man-plus-machine scenario, especially through the interactivity of AR, which would allow the visualization of the decision maps and, thus, identify whether the classifier is confidently producing a prediction (Strodthoff et al. 2020).

Finally, the current price and limited availability of high-end AR devices might represent an obstacle for the translation of the developed pipeline into daily clinical routine. Nonetheless, the developed pipeline could be easily ported to any portable device with a camera such as a mobile phone. This will be part of future work together with a clinical validation of the tools with real printed ECG recordings from our collaborators.
References


López Farré A, Macaya Miguel C. 2009. Libro de la salud Cardiovascular del Hospital Clínico San Carlos y la Fundación BBVA. Fundación BBVA.


