







The Deep-Match Framework: R-Peak Detection in Ear-ECG

Harry J. Davies , *Member, IEEE*, Ghena Hammour , Marek Zylinski , Amir Nassibi , *Member, IEEE*, Ljubiša Stanković , *Fellow, IEEE*, and Danilo P. Mandić , *Fellow, IEEE*

Abstract—The Ear-ECG provides a continuous Lead I like electrocardiogram (ECG) by measuring the potential difference related to heart activity by electrodes which are embedded within earphones. However, the significant increase in wearability and comfort enabled by Ear-ECG is often accompanied by a degradation in signal quality - an obstacle that is shared by the majority of wearable technologies. We aim to resolve this issue by introducing a Deep Matched Filter (Deep-MF) for the highly accurate detection of R-peaks in wearable ECG, thus enhancing the utility of Ear-ECG in real-world scenarios. The Deep-MF consists of an encoder stage, partially initialised with an ECG template, and an R-peak classifier stage. Through its operation as a Matched Filter, the encoder searches for matches with an ECG template in the input signal, prior to filtering these matches with the subsequent convolutional layers and selecting peaks corresponding to the ground-truth ECG. The latent representation of R-peak information is then fed into a R-peak classifier, of which the output provides precise R-peak locations. The proposed Deep Matched Filter is evaluated using leave-one-subject-out cross-validation over 36 subjects with an age range of 18-75, with the Deep-MF outperforming existing algorithms for R-peak detection in noisy ECG. The Deep-MF achieves a median R-peak recall of 94.9% and a median precision of 91.2% across subjects when evaluated with leave-one-subject-out cross validation. Overall, this Deep-Match framework serves as a valuable step forward for the real-world functionality of Ear-ECG and, through its interpretable operation, the acceptance of deep learning models in e-Health.

Index Terms—Convolutional neural networks, deep learning, electrocardiography, matched filters, wearable health monitoring systems.

I. INTRODUCTION

RECENT advancements in Hearables serve to disrupt the e-Health market through the provision of continuous monitoring of mental state and vital signs from the ear [1]. Of

the different sensing modalities within Hearables, one of the most notable is the Ear-ECG, which provides continuous Lead I electrocardiogram (ECG) [2], the measurement of the electrical activity of the heart, through the potential difference between two in-ear electrodes on opposite sides of the head [3], [4]. The precise heart rate information from ECG can be used to monitor stress through heart rate variability metrics [5], [6], [7] and the detection of irregular heart rhythms (arrhythmia) [8]. Whilst a chest patch would be preferable to screen for specific heart conditions, the option of using Ear-ECG opens the opportunity of detecting irregular heart rhythms in manifold new scenarios. Moreover, it is feasible that Ear-ECG could provide heart rate in cases where ear-EEG is already being utilised, such as the monitoring of sleep [9] and the monitoring of driver fatigue [10]. However, the considerable gain in comfort and wearability afforded by an in-ear sensor, compared to electrodes on the chest, comes at the expense of the signal to noise ratio (SNR). This is because the cardiac-induced potential difference is far smaller across the head than it is from other recording locations such as the chest or arms [3]. Indeed, the potential difference across the heart is often as much as 2 orders of magnitude lower from the ear than it is at the chest [11]. Moreover, the Ear-ECG commonly contains other signals comparable in amplitude, such as electrical activity generated by eye movements, known as electrooculography (EOG) [12], [13], and electrical signals generated by neuronal activity in the brain, known as electroencephalography (EEG) [9], [14], [15], [16], [17]. In order to best exploit the benefits of Ear-ECG, algorithms need to be able to detect the presence of ECG waveform across challenging range of signal qualities, and to correctly distinguish the peaks in ECG (R-peaks) from peaks that may occur due to artefacts or other electrical activity.

Given the well defined structure of ECG, it is natural to ask whether a matched filter [18], the process of finding a template and its position in a noisy signal through cross-correlation, can be used to detect R-peaks in wearable ECG [19], [20]. This has been demonstrated previously through the combination of matched filter and Hilbert transform [21], which was shown to outperform the commonly used Pan-Tompkins algorithm [22] for R-peak detection. Recent work on the interpretability of convolutional neural networks (CNNs) has demonstrated that at a fundamental pattern recognition level, a CNN operates in the same way as a matched filter, by performing convolution between a learned template kernel and an input signal or image and exploiting the correlation between the two [23].

Manuscript received 18 June 2023; revised 13 December 2023; accepted 18 January 2024. Date of publication 29 January 2024; date of current version 20 June 2024. This work was supported in part by the USSOCOM MARVELS grant and in part by the Dementia Research Institute at Imperial College London. (*Corresponding author: Harry J. Davies.*)

Harry J. Davies is with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ London, U.K. (e-mail: harry.davies14@imperial.ac.uk).

Ghena Hammour, Marek Zylinski, Amir Nassibi, and Danilo P. Mandić are with the Department of Electrical and Electronic Engineering, Imperial College London, U.K.

Ljubiša Stanković is with the University of Montenegro, Montenegro. Digital Object Identifier 10.1109/TBME.2024.3359752



Fig. 1. Ear-ECG earpiece. Left: The placement of one of the ear electrodes within the ear canal. Right: A labelled prototype Ear-ECG device, consisting of a foam ear-plug, a cloth electrode and an ear-hook to stabilise the ear-piece within the ear canal. A gold-cup electrode was placed on the forehead to provide the ground.

This was further verified through the MNIST handwriting data set, whereby trained kernels converged to resemble different numbers [24]. Given the clear benefits of using matched filtering to detect R-peaks in noisy ECG, and the theoretical link between CNNs and matched filtering [23], it is hypothesised that a learned convolutional matched filter could be leveraged to provide superior results for R-peak detection, whilst remaining fully interpretable in its operation.

To this end, we implement a deep convolutional neural network based matched filter for the efficient and accurate detection of R-peaks in Ear-ECG with poor signal to noise ratio. The trained model, whilst demonstrating exceptional performance over existing methods, has the benefit of full interpretability through the lens of matched filters, with kernel weights that exploit and amplify aspects of the ECG pattern.

II. METHODS

A. Hardware and Data

Simultaneous Ear-ECG and either arm-ECG or chest-ECG (resembling Lead I) was measured from 36 subjects, with an age range of 18–75. There was a minimum of 2 minutes of data recorded from each subject, with the majority of subjects having 5 minutes of ECG data. Recordings took place when subjects were still or during sleep to minimise the impact of motion artefacts, but it should be noted that motion artefacts were still present in the data, albeit rare, and not excluded from our analysis. Moreover, since many consumers wear earbuds when sedentary, examining results only on subjects who were still does not detract from the real world applicability of this technology. In 34 of the subjects, who were awake and still, the Ear-ECG was recorded with two earpieces across the head with a ground electrode placed on the forehead. In two of the subjects, who were sleeping, the Ear-ECG signal was from a single ear electrode which was referenced to the contra-lateral mastoid. The Ear-ECG earpiece, shown in Fig. 1, consisted of a foam earpiece with a cloth electrode, and electrode gel was used to reduce the impedance between the electrodes and the skin of

the ear canal. The data set contained data recorded from two different amplifiers, namely the BrainAmp from BrainVision (North Carolina, USA) in the subjects that were awake, and the Somno HD amplifier from Somnomedics (Randersacker, Germany) in the subjects that were sleeping. Utilising data from two different electrode configurations and amplifiers served to confirm that our model was not over-fitting to a specific recording paradigm. The recordings were performed under the IC ethics committee approval JRCO 20IC6414. All subjects gave full informed consent.

The Ear-ECG was down-sampled from 500 Hz to 250 Hz, and pre-filtered with three separate configurations to provide 3 input channels to the model. The first channel was a band-pass filtered Ear-ECG between 1 Hz and 45 Hz which aimed to reduce higher frequency noise whilst preserving the crucial information in the ECG. The second channel was a band-pass filtered Ear-ECG between 1 and 5 Hz, which removed high frequency noise and the QRS complex from the ECG, but retained information on the P and T waves. The third channel was a high-pass filtered Ear-ECG filtered with a cut-off frequency of 1 Hz. This preserved all of the higher frequency detail present in the ECG, but also retained high frequency noise such as electrical interference at 50 Hz. To segment the data, a sliding window with a length of two seconds (500 samples) was implemented with a shift of 0.4 seconds (100 samples). This resulted in a total of 26,564 segments across all subjects. Two seconds was chosen as the segment length so that inputs would always have an ECG waveform, and usually have upwards of two ECG waveforms.

B. Matched Filter Theory

Matched filtering is a signal processing technique that originated in radar applications [25]. It involves performing a cross correlation between the noisy signal $x(n)$, which contains a signal of interest and noise, and the known signal pattern $h(k)$, resulting in the output signal $y(n)$ as follows:

$$y(n) = \sum_k h(k)x(n+k) \quad (1)$$

When the defined pattern overlaps with itself in the noisy signal, there is a “match” which results in a peak in the output. The matched filter therefore maximises the signal to noise ratio of the output. In our case, when searching for peaks in noisy ECG, the matched filter performs the cross correlation between a template ECG and the noisy Ear-ECG.

A previous implementation of the matched filter in noisy R-peak detection is a combination of matched filtering and the Hilbert transform (MF-HT) introduced by Chanwimalueang et al. [21]. This algorithm requires a user to manually select the QRS complex in the data, which is then used as the matched filter template. The data is windowed for a duration after the current R-peak, with the start at 20% of the mean RR interval and the end at 150%. This window is then matched filtered, before the Hilbert transform is applied. The magnitude of the Hilbert transform gives the positive envelope, on which peak detection is performed. Potential peaks are windowed to the same length of the original QRS template, and cross correlation is performed with the original QRS template to determine the best match. If

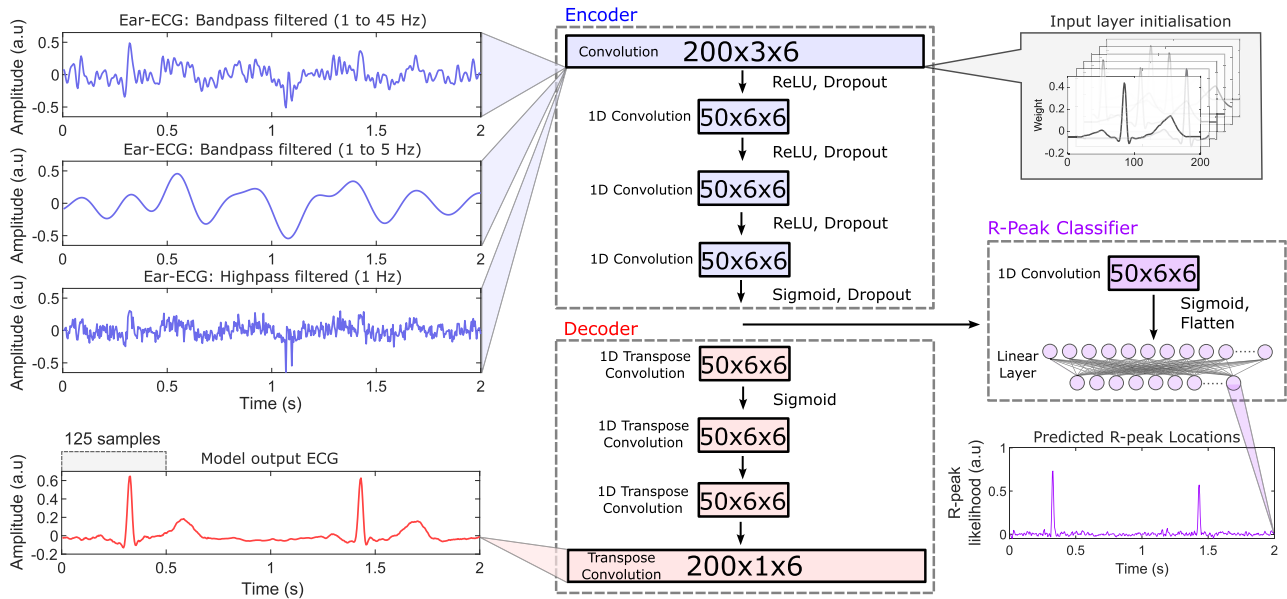


Fig. 2. Overview of the proposed Deep Matched Filter (Deep-MF) architecture. The three input channels to the model (top left, blue), including Ear-ECG band-pass filtered between 1 and 45 Hz, Ear-ECG band-pass filtered between 1 and 5 Hz, and Ear-ECG high-pass filtered with a cut-off frequency of 1 Hz. These channels serve as inputs to an encoder module which serves as a matched filter (middle, blue). The encoder, constructed of 1D convolutional layers, consists of a matched filter layer with kernels of length 200 (0.8 seconds) which serve to detect ECG patterns in the input, and three subsequent “refinement” layers with kernels of length 50, to determine which matches are true. A subsection of the weights of the matched filter layer are initialised with a shifted ECG template (top right, grey). The encoder is accompanied by a decoder (middle, red), consisting of 1D transpose convolutional layers, which upsample the output of the encoder into an output which resembles an ECG waveform (bottom left, red). The decoder is essential for the training of the encoder. The final module is the R-peak classifier (right, purple), which takes the output of the encoder and uses it to predict the position of the R-peak. The R-peak classifier consists of a single 1D convolution layer, and a linear layer.

the newly determined RR interval of the best match is outside the desired range, the RR interval is calculated for all potential peaks and the peak with the closest RR interval to the previous one is selected. If it is within the desired range, then the best match is selected as the R-peak. This process then repeats for all subsequent R-peaks.

C. Deep Matched Filter Model

The deep matched filter architecture, developed in PyTorch [26] and shown in Fig. 2, consists of two main parts. Firstly, an encoder-decoder module, which aims to extract shared information between the input and a training reference by condensing the information from the input that is most predictive of the output into a latent representation [27]. In this case, the encoder-decoder was trained with arm-ECG as a reference and aims to encode the shared information between the Ear-ECG and the arm-ECG, before decoding this information into a waveform resembling that of the arm-ECG. The encoder, whilst similar in organisation to that of a denoiser, behaved as a matched filter by simply encoding the R-peak locations from the original Ear-ECG and no corresponding morphological information. It then used this encoded R-peak location and generated a learned generic ECG pattern in the same position. The decoder was thus bypassed, with a simple CNN based classifier which used the latent representation to predict the R-peak location.

The encoder, highlighted in blue in Fig. 2, consisted of 4 one-dimensional convolutional layers. In the first layer, there were 6 kernels associated with each input channel, to form 6 output channels. In all subsequent layers there were 6 kernels

associated with each of the 6 new inputs. In the first layer, a kernel size of 200 was chosen, corresponding to 0.8 seconds and representing a duration slightly longer than that of a full ECG segment for its use as a matched filter template. Moreover, the 6 kernels corresponding to the first band-pass filtered input channel (1 to 45 Hz) were initialised with a shifted ECG template which is highlighted in grey in Fig. 2. The subsequent layers in the encoder had a kernel size of 50, chosen to encompass the width of the resulting “match” peak from convolution between the input and the input layer. These layers served as refinement layers for the output of the first layer, in essence helping the model to increase the precision of the matched filter by deciding which matches were valid and which matches were not. The first 3 layers had a ReLU activation function and a dropout of 50%, and the fourth layer had a Sigmoid activation function and a dropout of 50%. The Sigmoid activation function was important for ensuring stability of the model during training, due to the bounded output property.

The decoder, highlighted in red in Fig. 2, contained 4 transpose convolutional layers which mirrored the corresponding one-dimensional convolutional layers of the encoder. In contrast to the encoder layers, there was no dropout applied and only a single Sigmoid activation function was applied between the first and second decoder layers. Moreover, there was a single output corresponding to a 2 s ECG trace. The encoder-decoder structure was trained to minimise mean squared error between the output and the reference ECG waveform. Importantly however, due to the encoder operating as a matched filter and the fact that there were only slight differences in the morphology of ECG waveform across subjects, the model minimised error

by detecting only the location of the ECG in the input, and upsampling this into a generic ECG waveform. Despite this, the training paradigm of using a decoder to replicate a full ECG waveform was necessary, as when the same structure was trained to replicate just the R-peaks it often failed to converge.

Given that the latent representation contained information on location of the ECG in the input, a simple classifier, highlighted in purple in Fig. 2, was then trained to take in the latent variables and output R-peak locations. This classifier consisted of a single layer 1D convolution, followed by a Sigmoid activation function and flattening, before finally being passed to a linear layer. For the training of this second model, the latent variables were extracted from the training inputs which were passed through the trained encoder. These latent variables were then used as inputs to the R-peak classifier which was trained against the corresponding R-peak locations from the reference ECG. The R-peak locations were calculated from the ECG reference using the MATLAB (ver. 2022b) function *findpeaks*, and both the location of the R-peak and the two neighbouring values were assigned a value of 1. Extending the window of the R-peak location from 1 to 3 in the training reference gave the model slightly more lenience in the shift of a peak, as without this step the model had a tendency to suppress peaks. The output of the classifier was trained to minimise mean squared error (MSE) against the corresponding binary R-peak array. Finally, averaging was performed on the 2-second output of the deep matched-filter, with a shift of 0.4 seconds. In a real-world setting it would be practical to implement the model with a rolling output, rather than waiting for each new 2-second window to pass. Moreover, if the ECG in the input was at the boundaries of the model and not a full waveform it would be difficult to detect. This issue is circumvented by using a rolling window by ensuring that every ECG waveform in the input is at some point close to the center of the input.

The encoder-decoder model was trained for 10 epochs and the R-peak classifier was trained for 15 epochs, with both numbers of epochs chosen purposely as to limit over-fitting. Both were trained with a batch size of 10 segments, and both the encoder-decoder and the R-peak classifier models were trained using leave-one-subject-out cross validation.

The path that the input signal takes through the combined model is highlighted in Fig. 3. Observe that in this test example, where multiple peaks are present in the input with only true R-peak, the output of the ECG template “matched filtering” layer results in 3 strong peaks. These peaks are then sifted through by the subsequent decoder layers to produce a correct peak in the latent representation - a process we refer to as “refinement”. This peak in the latent representation is then used by the R-peak classifier to determine the true R-peak location. It should be noted that in examples with a higher heart rate, the latent representation consists of multiple correct peaks.

D. Model Evaluation

The deep matched filter (Deep-MF) was evaluated against two state of the art models, namely a standard matched filter (MF) and the matched filter Hilbert transform algorithm

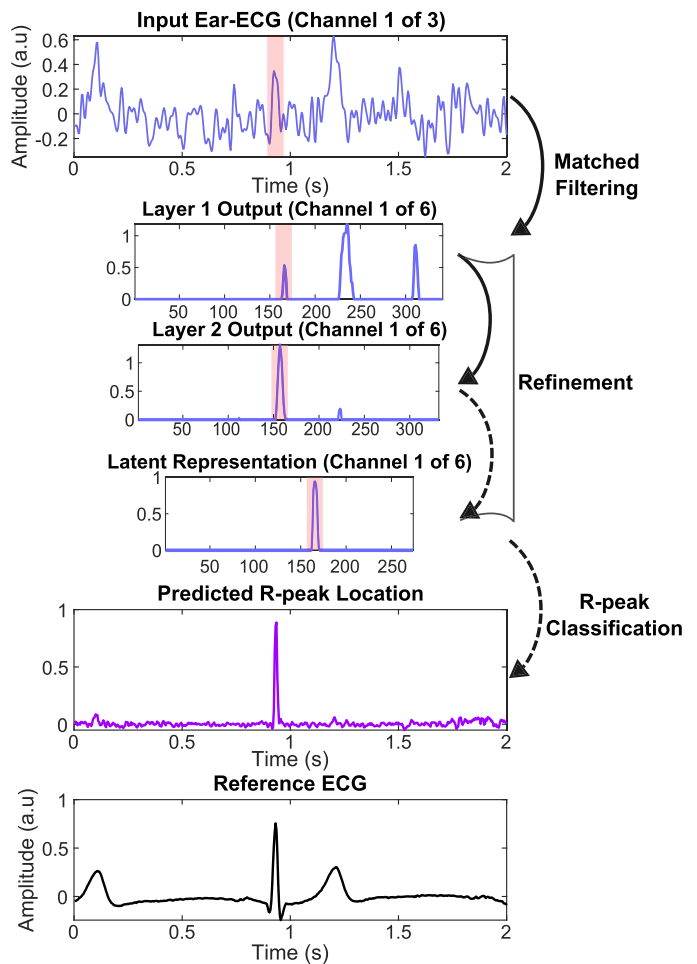


Fig. 3. Signal pathway through the proposed Deep Matched Filter (Deep-MF) for an example test input Ear-ECG trace (blue). The input first passes through the matched filter layer, resulting in the Layer 1 Output, with 3 potential matches. This initial output is then passed through the subsequent “refinement” layers, until a singular peak is present in the latent representation. In the matched filter and refinement stages, the true peak is highlighted with a shaded red box. This latent space is then passed through the R-peak classification phase, resulting in predicted R-peak location (purple). For the purposes of comparison, the ground truth ECG is displayed below in black.

(MF-HT) [21]. Both MF and MF-HT were implemented using the input channel that was band-pass filtered between 1 Hz and 45 Hz. For the outputs of the Deep-MF and MF, R-peaks were determined using the MATLAB function *findpeaks*, with a maximum peak width of 25 samples and a minimum peak distance of 12 samples. The determined R-peaks were then compared to the true R-peaks, which were also calculated using *findpeaks* on the reference ECG signal. If the predicted peak was within 40 ms of the true R-peak, it was considered a match. This condition was also applied to the output of the MF-HT.

The proposed Deep-MF and the MF were both evaluated in terms of R-peak recall (the proportion of the R-peaks in the reference signal that were correctly identified) and R-peak precision (the proportion of predicted R-peaks which were true R-peaks). An area under the curve (AUC) value corresponding to a precision-recall curve was calculated for both the proposed Deep-MF and the standard MF, by varying the minimum peak height threshold of the *findpeaks* function. This precision-recall

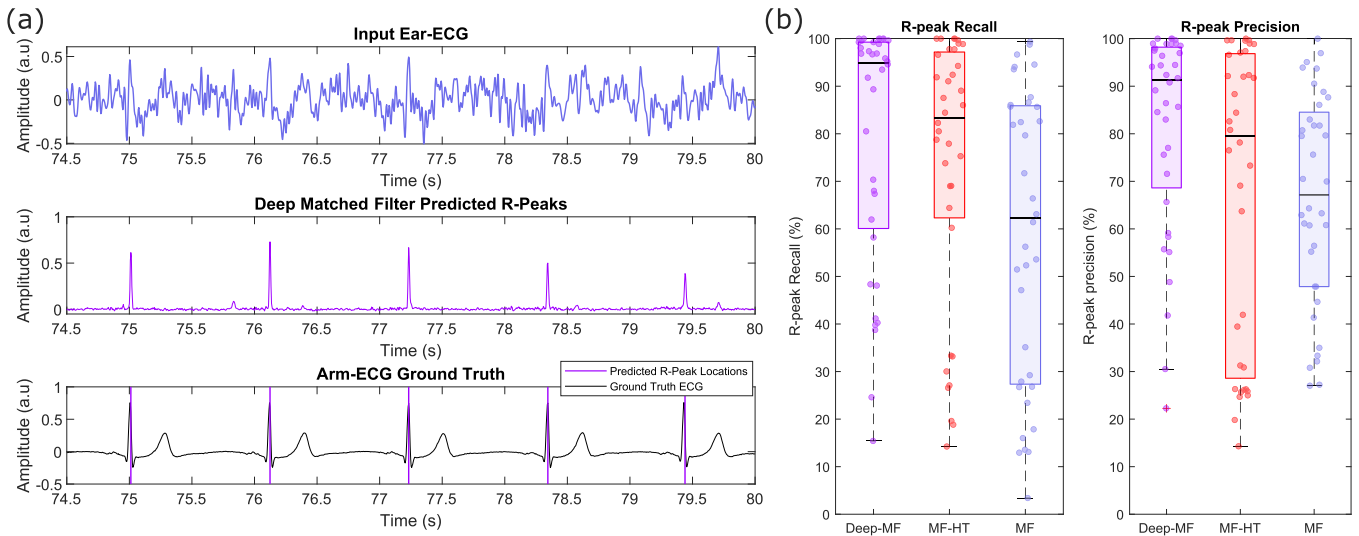


Fig. 4. Test results for the proposed Deep Matched Filter for R-peak detection in noisy Ear-ECG. (a) An example of Ear-ECG with a poor signal to noise ratio (blue) and the corresponding output R-peak locations of the Deep-MF (purple). Below is the ground truth Arm-ECG (black) with predicted R-peak locations overlaid in purple. Note that in the input there are several peaks which are stronger than the true R-peaks, particularly between 79 and 80 seconds. The Deep-MF correctly rejects these peaks and predicts the true R-peaks in the output. (b) Boxplots of R-peak Recall and Precision, across all subjects as a result of leave one subject out cross validation. Overlaid in each case is a swarm plot, showing the performance metrics for each subject individually. The results of the proposed Deep-MF filter (purple) are compared to the Matched Filter Hilbert Transform (MF-HT) (red) and the standard Matched-Filter (blue). In terms of R-peak recall, the percentage of peaks in the ground truth correctly identified by the model, the proposed Deep-MF achieves a median of 94.9%, compared with the MF-HT and MF where the respective median recalls are 83.4% and 62.3%. In terms of precision, the percentage of the peaks predicted by the model that are correct, the proposed Deep-MF achieves a median of 91.2%, compared with the MF-HT and MF which achieved the respective median precisions of 79.5% and 67%.

curve was generated with the median precision and recall values across all subjects. For the MF-HT algorithm, it was not possible to vary sensitivity in this way, and thus the implementation was compared Deep-MF and the standard MF with fixed threshold parameters that produced a good balance of recall and precision, with values of 0.11 in the case of Deep-MF and 0.90 in the case of MF. Note that the large difference in threshold used between the Deep-MF and the standard MF stemmed from the fact that the outputs of the Deep-MF were scaled and were thus lower in amplitude than the standard MF. The Deep-MF, standard MF and MF-HT were all compared again through performance in recall and precision, in the form of boxplots of these values across all 36 subjects.

For a more complete overview of model performance, precision and recall were further assessed against the signal quality of the model inputs. Signal quality was determined as the Pearson correlation coefficient between the 1–45 Hz bandpass filtered Ear-ECG input and the reference ECG signal. The distribution of signal quality across all 36 subjects was analysed, as well as the correlation between signal quality and model performance metrics for each algorithm.

The effects of initialisation with an ECG template (shown in grey in Fig. 2) were also evaluated and contrasted against random initialisation, both in terms of performance and interpretability. To evaluate the performance impact of initialisation, the mean absolute test error of the encoder-decoder model was calculated at regular intervals during training for both random initialisation and ECG template initialisation. Similarly, AUC values for R-peak recall-precision curves were calculated for both random initialisation and ECG template initialisation. For the purposes

of interpretability, the converged kernel weights after training with ECG template initialisation were examined visually and compared to the initialised values, with a focus on the P, Q, R, S, and T portions of the ECG to determine which aspects of the ECG were valuable to the model for detecting ECG in the input.

III. RESULTS AND DISCUSSION

The deep matched filter achieved a median R-peak detection recall of 94.9% in the Ear-ECG of unseen subjects, with an interquartile range of 60.1% to 99.3%. The Deep-MF had a corresponding median precision of 91.2% with an interquartile range of 68.6% to 98.2%. The high recall and precision of the Deep-MF is reinforced by an example Deep-MF model output shown in Fig. 4(a), alongside the ground truth ECG and the input Ear-ECG. It can be observed in this example that even in a scenario with a poor signal-to-noise ratio, whereby it is difficult to visually identify which peaks in the input belong to ECG, the Deep-MF correctly identified the correct peaks and excluded the incorrect peaks. On the same subject pool, the MF-HT achieves a median recall of 83.4% with an IQR of 62.3% to 97.2%. The MF-HT had a corresponding median precision of 79.5% (IQR 28.6% to 96.8%). These results are comparable to the original implementation of MF-HT on noisy ECG by Chanwimalueang et al. [21], in which the algorithm achieved a recall of 83.1% and a precision of 86.8%. The increase in precision achieved by the deep-MF over MF-HT was statistically significant when analysed through one-way ANOVA ($P = 0.03$) but the improvements in recall were not.

The results of the Deep-MF and MF-HT were also compared with the standard MF which had a median recall of 62.3% (IQR 27.3% to 85.9%) and a median precision of 67.1% (IQR of 47.9% to 84.5). The improvement in performance of the deep-MF over the standard MF was statically significant both in terms of recall ($P. = 0.003$) and precision ($P. = 0.005$). The full results for the comparison of recall and precision between the Deep-MF, MF-HT and standard MF are shown in boxplots in Fig. 4(b) with the Deep-MF plotted in purple, the MF-HT in red, and the standard MF in blue. Whilst the overall distribution of recall was comparable between the Deep-MF and MF-HT, with both models having a similar interquartile range, Deep-MF performed far better in terms of precision. This is likely due to the advantages of the Deep-MF related to having multiple refinement layers, in which false peaks could be discarded. It is important to note that the MF-HT relies on manual input to select a matched filter template from the input signal, which explains the improvements in recall over the standard-MF in which a fixed template was used across all subjects. Moreover, the MF-HT also has further conditions on determining true R-peaks, based on a balance between the correlation between the template and the input and a deviation from the mean RR interval. These conditions explain why the median precision of the MF-HT was also higher than for the standard MF.

The interquartile ranges of both recall and precision were large for all three algorithms, and this was primarily a function of the large variability in signal quality across subjects. The signal quality (determined as the Pearson correlation between the input Ear-ECG and reference ECG) had a median correlation of 0.39, with an interquartile range of 0.24 to 0.49. For reference of what these values mean in practice, the Ear-ECG of the subject displayed in Fig. 4(a) had a signal quality of 0.40. Indeed, this highly variable signal quality was either moderately, strongly or very strongly correlated with each performance metric for each type of matched filter algorithm employed. For the Deep-MF, there was a correlation of 0.68 between signal quality and recall ($P. = 6 \times 10^{-6}$), and a correlation of 0.63 between signal quality and precision ($P. = 4 \times 10^{-5}$). For MFHT, the correlations between performance and signal quality were 0.60 ($P. = 1 \times 10^{-4}$) and 0.49 ($P. = 2 \times 10^{-3}$) for recall and precision, respectively. For the standard MF, the signal quality correlations were 0.82 ($P. = 1 \times 10^{-9}$) and 0.58 ($P. = 2 \times 10^{-4}$) for recall and precision, respectively. Moreover, when only the performance on the worst quarter of signal qualities was considered, there was no statistically significant difference in performance between different algorithms. This is in contrast to the statistically significant improvement in performance provided by the Deep-MF across all subjects. In this work, gel was used to improve signal quality, which would not be utilised in consumer grade devices. The lack of improvement in performance on poor data indicates that for the Ear-ECG to be reliable in real world devices, advancements must be made in the sensor hardware and thus the consistency of the acquired in-ear ECG signals. Recent promising advancements in dry electrode technology exist which have the potential to overcome this hurdle [28], [29].

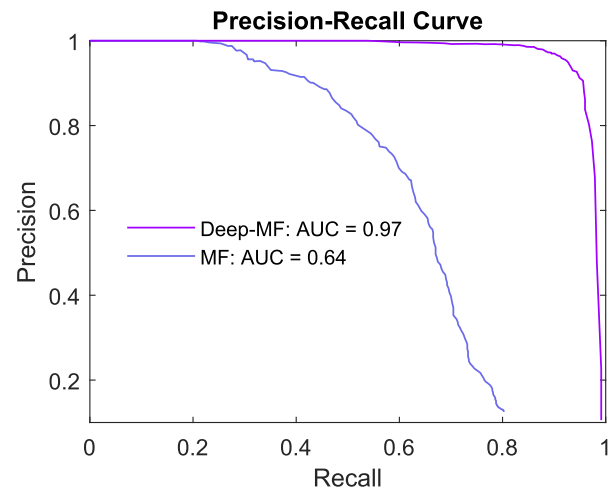


Fig. 5. Precision-recall curves for R-peak detection, for the proposed Deep Matched Filter (Deep-MF) and the standard matched filter (MF). The Deep-MF (purple) achieves an area under the curve (AUC) value of 0.97, compared with the MF (blue) which achieves an AUC of 0.64.

For both the Deep-MF and standard MF, recall and precision were evaluated across the full range of peak sensitivity threshold values to produce precision-recall curves taken from the median results across all subjects, as shown in Fig. 5. The Deep-MF achieved an area under the curve value of 0.97, compared to the standard MF which achieves an AUC of 0.64. Observe that precision values never drop below 0.1 due to the two fixed conditions of this *findpeaks* implementation, namely the maximum peak width and a minimum peak distance, which limited precision from dropping below this value.

It is important to note that the Deep-MF model is more complex than the MF-HT and the standard MF, which provides a more significant barrier to practical implementation. However, on the landscape of deep learning implementations, the Deep-MF is a relatively small model, with an implementation consisting of only 6 total layers, of which 5 are convolutional layers. Moreover, the total number of convolutional kernels is only 162, making the Deep-MF very computationally cheap to implement. This small architecture, grounded in the matched filter principle and combined with the fact that the overall structure of ECG shares many features across participants, allowed us to achieve high accuracy with a relatively small subject pool. As is the case with the standard MF and MF-HT, the Deep-MF can operate in quasi real-time, with a rolling window of the previous 2 seconds of input data. Furthermore, the increase in model complexity of the Deep-MF is justified by vast improvements in performance, with a median increase in recall of 11.6% and a median increase in precision of 11.8% when compared to the MF-HT.

IV. INTERPRETABILITY

A major barrier to the more widespread adoption of deep learning techniques in digital health is the notion that the models themselves are black boxes that offer no interpretability to explain the predictions they make. In this work, we employ the

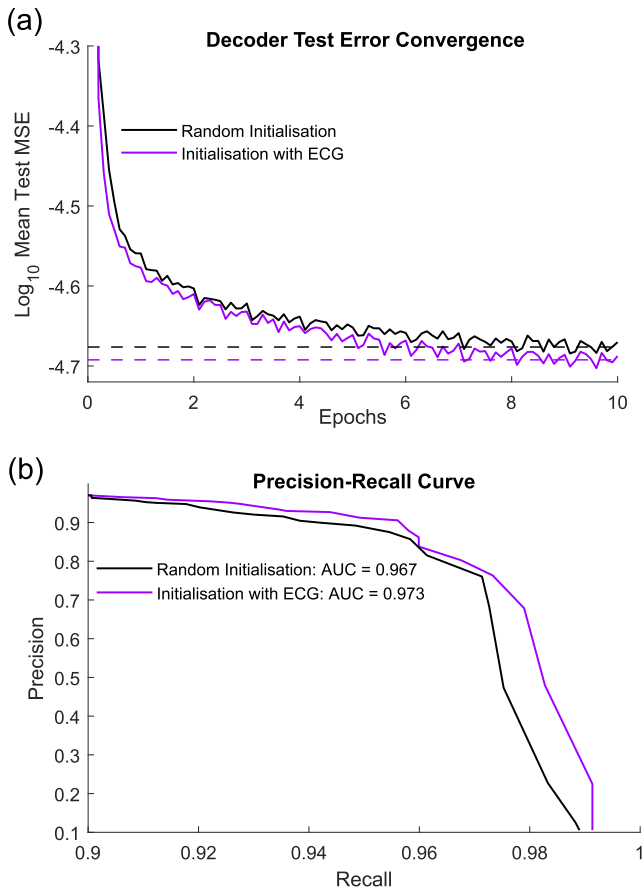


Fig. 6. Effects on model performance of the random initialisation of kernel weights (black), against partial initialisation of the input "Matched Filter" layer with an electrocardiogram template (purple). (a) The Log_{10} of the test error convergence of the encoder-decoder module during the training process, with a dotted line representing the mean test error of the last half of the final training epoch. (b) Precision-Recall curves for R-peak detection of the Deep-MF, shown to have an area under the curve (AUC) value of 0.967 with random initialisation and an AUC value of 0.973 after initialisation with an ECG template.

concept developed by Stanković et al. [23] to show that this particular model for R-peak detection is fully interpretable as a multi-layer matched filter, a tool for detecting similarity in patterns between an input signal and a template of interest. A key aspect of this argument is the partial initialisation of the input layer of the Deep-MF to the pattern which it is trying to detect [24], namely the electrocardiogram. For rigour, it is important to also examine the same kernels after the training process, as if the network was to completely discard the ECG template then it could be assumed that the model did not find the information useful for minimisation of error and thus was not searching for ECG "matches" in the input signal.

In terms of model accuracy, the effects of partial initialisation of the input layer of the Deep-MF with the same ECG template as used in the standard matched filter were small. It is highlighted in Fig. 6(a) that initialisation with an ECG template provided a minor improvement in the mean squared error of the decoder output, and in Fig. 6(b) it is shown that this corresponded to a slight increase in precision-recall AUC from 0.967 to 0.973.

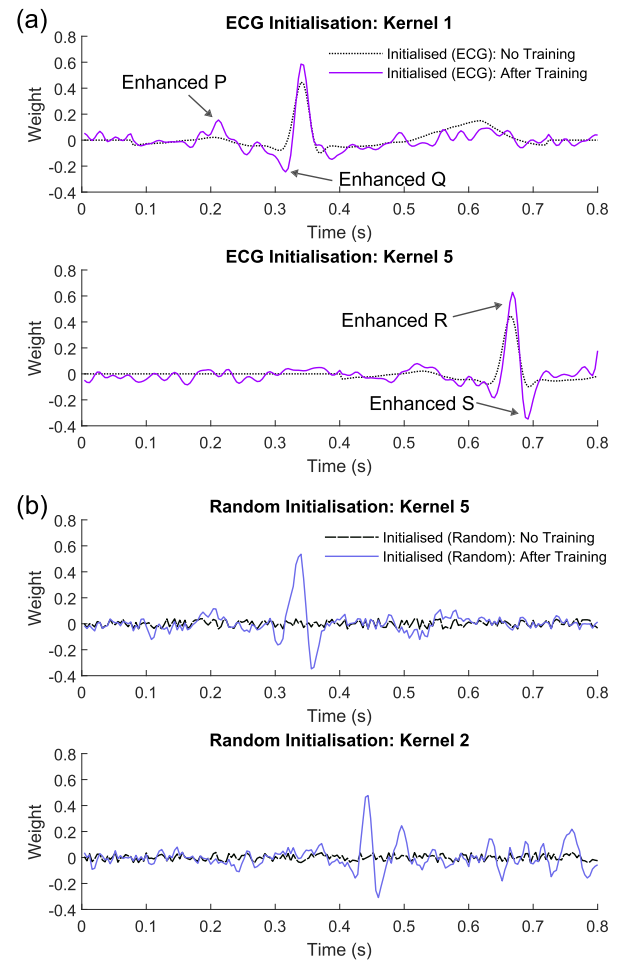


Fig. 7. Effects of training on kernels initialised with or without an electrocardiogram template. (a) The dashed trace in black represents the initialised ECG kernel weights before the training process, and the purple trace represents the kernel weights after the training process. Labelling highlights that across kernels, different aspects of the ECG are enhanced during the training process. In the 1st kernel (top), there is an enhancement of the P wave and of the Q portion of the QRS complex. In the 5th kernel there is an amplification of the S portion of the QRS complex. In all ECG initialised kernels there is an exaggeration of the R-peak that was provided by the ECG template. (b) The dashed trace in black represents the randomly initialised weights before the training process, and the blue trace represents the kernel weights after the training process.

Note that the precision-recall curves plotted in Fig. 6(b) are zoomed in the recall axis to exaggerate the difference between the random initialisation and the template initialisation. Whilst these improvements in performance are marginal, it does suggest that initialising the Deep-MF with an ECG provided the model with useful information that it did not otherwise learn from the training process.

When examining the effect of training on the initialised weights, as shown with two examples in Fig. 7(a), it is clear that the network retains aspects of the ECG templates as it deems them useful in minimising error. Notably, in the all kernels initialised with an ECG template, the R-peak information from the template was retained by the network and exploited with an increase in the weights at this location. Moreover, the network

goes further than the original template ECG and exaggerates aspects such as the P wave, and the Q and S parts of the QRS complex [30], [31] during the training process, showing that these aspects of the ECG are useful in distinguishing true R-peaks from other peaks in the input signal. This can be seen in Fig. 7(a) with the kernel 1 demonstrating an amplification in weights around the P wave and Q, and kernel 5 demonstrating an exaggeration of the R and S components of the ECG. It is further shown, through the convergence of the randomly initialised weights in Fig. 7(b), that the model may also converge to an QRS matched filter on its own (kernel 5) but that this result is not necessarily guaranteed (kernel 2) without initialisation to an ECG template.

V. CONCLUSION

We have introduced a novel Deep Matched Filter (Deep-MF) framework for the detection of R-peaks in wearable-ECG. The proposed Deep-MF has been evaluated on the Ear-ECG of 36 subjects, and has shown a marked improvement over existing matched filter based algorithms, both in terms of recall and precision. In parallel with demonstrating the proficiency of the Deep-MF at R-peak detection in scenarios with poor signal to noise ratio, it has been illustrated that this encoder-based model behaves precisely as a learned matched filter. It serves to detect ECG segments in the input, followed by several refinement layers which distinguish the true ECG matches from the false matches. This has been reinforced through partial initialisation of the model with an ECG template, whereby through the training process the model enhances physically relevant aspects of the ECG. Owing to its interpretability, the proposed Deep Matched Filter has been shown to greatly improve the practical utility of the Ear-ECG signal, whilst being transparent in its operation. It is our hope that physically grounded models such as the Deep Matched Filter may help to accelerate wide scale adoption of interpretable artificial intelligence in healthcare.

REFERENCES

- [1] V. Goverdovsky et al., "Hearables: Multimodal physiological in-ear sensing," *Sci. Rep.*, vol. 7, no. 1, Jul. 2017, Art. no. 6948.
- [2] N. Arora and B. Mishra, "Origins of ECG and evolution of automated DSP techniques: A review," *IEEE Access*, vol. 9, pp. 140853–140880, 2021.
- [3] W. von Rosenberg et al., "Hearables: Feasibility of recording cardiac rhythms from head and in-ear locations," *Roy. Soc. Open Sci.*, vol. 4, no. 11, 2017, Art. no. 171214.
- [4] W. Von Rosenberg et al., "Electrocardiogram apparatus and method," U.S. Patent US11,633,142, Apr. 25 2023.
- [5] T. Adjei et al., "The ClassA framework: HRV based assessment of SNS and PNS dynamics without LF-HF controversies," *Front. Physiol.*, vol. 10, no. 505, pp. 1–15, 2019.
- [6] H.-G. Kim et al., "Stress and heart rate variability: A meta-analysis and review of the literature," *Psychiatry Investigation*, vol. 15, no. 3, pp. 235–245, 2018.
- [7] L. Pecchia, P. Melillo, and M. Bracale, "Remote health monitoring of heart failure with data mining via CART method on HRV features," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 800–804, Mar. 2011.
- [8] G. Hammour et al., "Hearables: Feasibility and validation of in-ear electrocardiogram," in *Proc. IEEE 41st Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2019, pp. 5777–5780.
- [9] T. Nakamura et al., "Hearables: Automatic overnight sleep monitoring with standardized in-ear EEG sensor," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 1, pp. 203–212, Jan. 2020.
- [10] M. C. Yarici et al., "Hearables: Ear EEG based driver fatigue detection," 2023, *arXiv:2301.06406*.
- [11] M. Yarici et al., "Hearables: Feasibility of recording cardiac rhythms from single in-ear locations," *Roy. Soc. Open Sci.*, vol. 11, no. 1, 2024, Art. no. 221620.
- [12] M. A. Skoglund et al., "Comparing in-ear EOG for eye-movement estimation with eye-tracking: Accuracy, calibration, and speech comprehension," *Front. Neurosci.*, vol. 16, pp. 1–16, 2022.
- [13] H. Manabe, M. Fukumoto, and T. Yagi, "Direct gaze estimation based on nonlinearity of EOG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 6, pp. 1553–1562, Jun. 2015.
- [14] V. Goverdovsky et al., "In-ear EEG from viscoelastic generic earpieces: Robust and unobtrusive 24/7 Monitoring," *IEEE Sensors J.*, vol. 16, no. 1, pp. 271–277, Jan. 2016.
- [15] P. Kidmose et al., "A study of evoked potentials from Ear-EEG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2824–2830, Oct. 2013.
- [16] K. B. Mikkelsen et al., "Sleep monitoring using ear-centered setups: Investigating the influence from electrode configurations," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 5, pp. 1564–1572, May 2022.
- [17] C. B. Christensen et al., "Ear-EEG-Based objective hearing threshold estimation evaluated on normal hearing subjects," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 5, pp. 1026–1034, May 2018.
- [18] G. Turin, "An introduction to matched filters," *IRE Trans. Inf. Theory*, vol. 6, no. 3, pp. 311–329, 1960.
- [19] Q. Xue, Y. H. Hu, and W. Tompkins, "Neural-network-based adaptive matched filtering for QRS detection," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 4, pp. 317–329, Apr. 1992.
- [20] P. S. Hamilton and W. J. Tompkins, "Adaptive matched filtering for QRS detection," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, 1988, pp. 147–148.
- [21] T. Chanwimalueang, W. von Rosenberg, and D. P. Mandic, "Enabling R-peak detection in wearable ECG: Combining matched filtering and hilbert transform," in *Proc. IEEE Int. Conf. Digit. Signal Process.*, 2015, pp. 134–138.
- [22] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.
- [23] L. Stanković and D. Mandic, "Convolutional neural networks demystified: A matched filtering perspective-based tutorial," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 53, no. 6, pp. 3614–3628, Jun. 2023.
- [24] S. Li et al., "Demystifying CNNs for images by matched filters," 2022, *arXiv:2210.08521*.
- [25] P. Woodward, "Simple theory of radar reception," in *Probability and Information Theory with Applications to Radar*, 2nd ed. Oxford, U.K.: Pergamon, 1953, pp. 81–99.
- [26] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8024–8035.
- [27] H. J. Davies and D. P. Mandic, "Rapid extraction of respiratory waveforms from photoplethysmography: A deep corr-encoder approach," *Biomed. Signal Process. Control*, vol. 85, no. 104992, pp. 1–8, 2023.
- [28] R. Sasaki et al., "Novel dry EEG electrode with composite filler of PEDOT: PSS and carbon particles," in *Proc. IEEE 45th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2023, pp. 1–4.
- [29] S. Guler et al., "Ear electrocardiography with soft graphene textiles for wearable applications," *IEEE Sens. Lett.*, vol. 6, no. 9, Sep. 2022, Art. no. 6003004.
- [30] H. Abrishami et al., "P-QRS-T localization in ECG using deep learning," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform.*, 2018, pp. 210–213.
- [31] A. A. R. Bsoul et al., "Detection of P, QRS, and T components of ECG using wavelet transformation," in *Proc. IEEE ICME Int. Conf. Complex Med. Eng.*, 2009, pp. 1–6.