Automated ECG diagnostic P-wave analysis using wavelets

A. Diery a,c, D. Rowlands a,c, T.R.H. Cutmore a,b,d,*, D. James a,c,e

a Centre for Wireless Monitoring Applications, Griffith University, Brisbane, 4111, Queensland, Australia
b Applied Cognitive Neuroscience Research Centre, Australia
c School of Engineering, Griffith University, Brisbane, 4111, Queensland, Australia
d School of Psychology, Griffith University, Brisbane, 4111, Queensland, Australia
e Centre of Excellence for Sport Science, Queensland Academy of Sport, Brisbane, Queensland, Australia

Abstract

P-wave characteristics in the human ECG are an important source of information in the diagnosis of atrial conduction pathology. However, diagnosis by visual inspection is a difficult task since the P-wave is relatively small and noise masking is often present. This paper introduces novel wavelet characteristics derived from the continuous wavelet transform (CWT) which are shown to be potentially effective discriminators in an automated diagnostic process. Characteristics of the 12-lead ECG P-wave were derived using CWT and statistical methods. A normal control group and an abnormal (atrial conduction pathology) group were compared. The wavelet characteristics captured frequency, magnitude and variance components of the P-wave. The best individual characteristics (i.e. ones that significantly discriminated the groups) were entered into a linear discriminant analysis (LDA) for four different models: two-lead ECG, three-lead ECG, a derived three-lead ECG and a factor analysis solution consisting of wavelet characteristic loadings on the factors. A comparison was also made between wavelet characteristics derived from individual P-waves verses wavelet characteristics derived from a signal-averaged P-wave for each participant. These wavelet models were also compared to standard cardiological measures of duration, terminal force and duration divided by the PR segment. Results for the individual P-wave approach generally outperformed the standard cardiological measures and the signal-averaged P-wave approach. The best wavelet model on the basis of both classification performance and simplicity was the two-lead model that uses leads II and V1. It was concluded that the wavelet approach of automating classification is worth pursuing with larger samples to validate and extend the present study.

© 2011 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The electrocardiogram (ECG) is a recording of the surface potential created by the electrophysiological processes of the cardiac cycle and used diagnostically by cardiologists and general practitioners [1]. This paper focuses on automated analysis of the P-wave, which represents the initial pulse created by the sino-atrial node and its conduction path through the atria.

Abnormalities in the P-wave may be produced by structural changes, such as left atrial enlargement (LAE), or inter-atrial block (IAB) [2]. LAE has been significantly linked to a host of pathological conditions and complications such as long-term and permanent atrial fibrillation, atrial thrombus and embolic stroke [3,4]. IAB has been shown to be significantly prevalent...
in patients who suffer embolic strokes [5]. Atrial conditions, such as these, can produce significant alterations to the electrical signature of atrial activity. Fig. 1 gives examples of P-waves recorded on the standard three-lead ECG configuration (V5, aVF and V2) and shows one normal ECG (column a) and two ECGs with abnormal atrial conduction. (The P-wave can also be seen as the small pulse to the left of the R wave spike in Fig. 2). Over the past 4 decades, several ECG indicators of LAE have been derived and used in clinical practice. However, in recent years these indicators have been shown to be unreliable and nonspecific to LAE [6–8] and IAB [9], underlining the need to define better methods for identifying and classifying atrial abnormalities from ECG recordings. Since ECG indicators of LAE are still used in everyday practice so that patients can be sent for more extensive testing or treatment, there remains the need to refine ECG analysis (for example with innovation in signal processing) to maximize the potential diagnostic information they may provide [10].

The dispersion of atrial conduction and the use of P-wave signal-averaged ECG have been shown to be reliable clinical predictors after surgery, forewarning the onset of arrhythmic atrial conditions [3,11–13]. As automated analysis of the ECG has evolved from experimental development into clinical practice, time–frequency analysis techniques have provided a significant addition to the analytic tools used by cardiologists [14].

Various approaches have been applied to classification of the ECG and specifically the P-wave. Several classic methods of system analysis have been applied to morphological classification of the P-wave. Both artificial neural networks [15], and system modelling [16] have been shown to be superior over traditional frequency domain and signal-averaged ECG (SAECG) methods (generally achieving an accuracy of about 85%). There are good reasons for examining the effectiveness of a wavelet-LDA P-wave classification system. First, while the use of wavelets for analysis and classification of biomedical signals, including some components of the ECG, are well documented [17,18], wavelet analysis specifically of the P-wave has not received much study. Wavelets offer an important information-rich parameterization method for data reduction of the ECG time-series. Secondly, neural networks functionally different from LDA (i.e. those with a hidden layer) require large samples due to the large number of parameters to be estimated. This is often not practical. Third, LDA is relatively assumption free, unlike model-based approaches; hence the wavelet-LDA approach described here offers parsimony with the potential for wide applicability.

The present study analysed P-waves from 12-lead ECG using wavelet decomposition. Characterization of the P-wave was performed using the wavelet energy spectrum, followed by statistical procedures that used these characteristics to discriminate between normal and abnormal groups. Since the approach taken required continuous frequency information across the spectrum the continuous wavelet transform (CWT) was used. The discrete version, the DWT, does not provide such detailed frequency information [19].

This study compared the following four methods to determine the best automated P-wave diagnostic classification inputs.

1. Three ECG lead method (aVF, V2, V5)
2. Two ECG lead method (II, VI)
3. Frank lead method (X,Y,Z) derived from the 12-lead ECG [20]
4. Factor analysis method.

Fig. 1 – P-waves at V5, aVF and V2 electrodes. (a) P-wave from a control participant; (b) and (c) show P-waves from two different participants exhibiting bifid atrial conduction. Signals have been low-pass filtered from 35 Hz.
2. Methods

2.1. Experimental data

Patient ECG was collected by a local cardiology clinic which had confirmed the diagnosis of abnormal left atrial conduc-
tion using a number of sources including history, clinical interview, ECG inspection and echocardiogram. This abnormal group \( (n = 32, \text{ age } 77.6 \pm 10.9 \text{ years}) \) included four subjects with possible inferior infarction, eight subjects with displayed sinus bradycardia, eight subjects with 1st degree AV block, and one subject with bi-atrial hypertrophy. Control group subjects \( (n = 7, \text{ age } 56.5 \pm 4.5 \text{ years}) \) which were screened by practising cardiologists, required them to be over 50 years of age with no history of heart disease or hypertension. The study received clearance by the Griffith University Ethics Council prior to commencement. The results from the control group should extrapolate reasonably well if a more aged normal group were used since IAB incidence increases from 32% to 50% between the ages of 40 and 50 but remains at 59% after the age of 60 [21]. Furthermore, discriminant analysis does not require equal or similar sample sizes [22].

All data were recorded using the standard 12-lead ECG configuration at 1000 Hz then down sampled with CardioView\textsuperscript{TM} [23] to 500 Hz and 3 dB band-pass filtered at 0.05–175 Hz. The ECG data was recorded in 10 s epochs as per clinical practice and this data was used in all of the analyses below. For half the abnormal group sample and for the control group there was 60 s of ECG which was only used in illustrating P-wave classification variability (see last paragraph of Section 3).

2.2. P-wave extraction

P-waves were identified by a standard approach involving automated multi-scale event detection to determine the location of the QRS complex as a landmark to search for the P-wave as shown in Fig. 2 [24]. Pre-processing of the ECG involved the removal of baseline drift using a 4th order low-pass Butterworth filter with a 3 dB cut-off at 0.5 Hz.

The QRS complexes were isolated by wavelet decomposition and applying a hard threshold to several selected scales of wavelet coefficients to emphasise the high frequency components [25]. A Mexican hat basis function was employed in a continuous wavelet transform to produce coefficients at scales with corresponding central frequencies of 25, 12.5, 8.93, and 6.94 Hz. This technique discriminates between QRS complexes and broader T waves which are produced from the same muscle mass and also have high amplitudes. A search window of 50–250 ms prior to each QRS complex was used to identify the P-wave peak. This window was found to be sufficient to account for the normal PR interval duration of 140–210 ms and the normal maximum Q-wave duration of 30 ms [1]. A second reconstruction of the ECG time series was computed dropping the high frequency (25 Hz) component to produce a P-wave with a clear minima associated with its onset and offset. The onset and offset values defined each individual P-wave for characterization and analysis. The duration of the P-wave was derived from the onset and offset of lead II which typically produces the longest duration over all the 12 leads. [1].

2.3. P-wave characterisation

Two approaches were used to characterise the P-wave. The first approach involved a wavelet decomposition of each individual P-wave followed by the computation of six quantities to capture the main features. The second approach used a signal average of the P-waves which was then processed using the same wavelet decomposition procedure. The signal-averaged approach has the possible advantage of improving the signal to noise ratio of the P-wave although with variability in the

Fig. 2 – This figure shows the signal processing steps involved in identification and extraction of P-waves from the raw time series. From this a database of 902 P-waves was compiled and submitted to either individual or beat averaged P-wave analysis.
individual P-waves it could also result in poorer performance due to smearing. These two approaches do not involve commutative processes and therefore differences in performance could be expected.

The chosen continuous wavelet transform (CWT) was the 2nd order derivative of the Gaussian function. This has been previously found to provide the highest correlation with the healthy P-wave [26].

The wavelet distribution was used to derive the energy, frequency and stability parameters. In deciding on the frequency bands to investigate, this study was guided by previous work [26] that found frequencies in the band 2.5–13.5 Hz to be highly correlated to the overall morphology of the P-wave. Frequency bands lower than 2.5 Hz were observed to be significantly affected by adjacent QRS complexes and were therefore discarded. This truncated distribution was cut into quartiles and two measures of wavelet parameter stability (IQR, QV) were derived from the quartiles since it has been shown that abnormal P-waves evidence variability [27]. All metrics on wavelet scale were converted to Hz for more transparent exposition of the processing.

The six-wavelet characterizations of the P-wave used in this study are defined below:

1. The measure for total energy was defined as the energy contained within all scales for the duration of the P-wave. Hence, the total energy summed over wavelet coefficients C(a,b) was:

\[
E_{vi}(a) = \frac{1}{a^2} \sum_{b=p_i}^{p_i+1} |C \{ a, b \}|^2
\]

2. The peak frequency (Hz) that corresponded to the maximum value of the wavelet transform coefficient’s highest frequency and stability parameters. In deciding on the frequency bands lower than 2.5 Hz were observed to be significantly affected by adjacent QRS complexes and were therefore discarded. This truncated distribution was cut into quartiles and two measures of wavelet parameter stability (IQR, QV) were derived from the quartiles since it has been shown that abnormal P-waves evidence variability [27]. All metrics on wavelet scale were converted to Hz for more transparent exposition of the processing.

3. The peak value of this wavelet scale (frequency)

4. The median frequency (Q2) of the wavelet distribution

5. Inter-quartile range (IQR)

\[
IQR = Q3 - Q1
\]

6. The normalized quartile variation (QV)

\[
QV = \frac{Q3 - Q1}{Q3 + Q1}
\]

The inter-quartile range describes the dispersion of the central 50% of the total energy of the P-wave across the wavelet frequency spectrum. The QV expresses the IQR as a percentage of the whole range of the energy distribution.

For individual P-wave analysis, the above algorithm was applied to each P-wave for a given lead, forming a matrix consisting of the number of P-wave in each recording, each with six columns (the parameters). Each column of parameters was reduced to its median value. For the averaged P-wave approach the algorithm was passed over the averaged waveform for each subject producing a single value for each parameter from the selected leads.

Four methods were used, each defined by a set of dimensions. These dimensions were either simply lead configuration (Methods 1 and 2), the Frank derived lead configuration (Method 3) or computed using factor analysis (Method 4) (see Fig. 3).

2.3.1. Two- and three-lead configurations

The two-lead configuration used lead II and V1 since this is the set most often used by cardiologists [1]. Lead II is oriented on the long axis of the heart and provides the longest P-wave duration. Lead V1 best differentiates the two atria.

The three-lead configuration used was aVF, V2, and V5, since this roughly corresponds to the axes of the heart [29]. Relative to the Frank lead configuration, lead V5 corresponds roughly to the X direction, lead aVF to the Y direction and lead V2 to the Z direction (but opposite in polarity).

Fig. 4 illustrates the three P-wave analysis methods on lead sets: (1) aVF, V2, V5; (2) II and V1; (3) Frank leads X, Y and Z. The Inverse Dower transform [30] was used to obtain the Frank X–Y–Z leads prior to wavelet decomposition. For each lead set, the six characteristics from each P-wave (\(E_{vi}(a)\), peak frequency, peak value, Q2, IQR, QV) were extracted as described in Section 2.3. The median of these characteristics for a particular subject formed the single feature vector for that subject. The feature vector matrix is given in Eq. (4).

\[
X^{lead1.lead2} = \begin{bmatrix}
V_{lead1}^{1} & V_{lead2}^{1} \\
V_{lead1}^{2} & V_{lead2}^{2} \\
V_{lead1}^{3} & V_{lead2}^{3} \\
V_{lead1}^{4} & V_{lead2}^{4} \\
V_{lead1}^{5} & V_{lead2}^{5} \\
V_{lead1}^{6} & V_{lead2}^{6}
\end{bmatrix}
\]

\[
X^{lead1.lead2.lead3} = \begin{bmatrix}
V_{lead1}^{1} & V_{lead2}^{1} & V_{lead3}^{1} \\
V_{lead1}^{2} & V_{lead2}^{2} & V_{lead3}^{2} \\
V_{lead1}^{3} & V_{lead2}^{3} & V_{lead3}^{3} \\
V_{lead1}^{4} & V_{lead2}^{4} & V_{lead3}^{4} \\
V_{lead1}^{5} & V_{lead2}^{5} & V_{lead3}^{5} \\
V_{lead1}^{6} & V_{lead2}^{6} & V_{lead3}^{6}
\end{bmatrix}
\]

Where

\[
V_{lead}^{n} = \{ E_{vi}(a), F_{peak}, V_{peak}, Q2, IQR, QV \}
\]

and \(n\) is the element number for each characteristic. Since the sample size was modest, parameter reduction prior to LDA was performed for each of the two-lead and three-lead models by selecting the best six discriminating elements from their respective matrices. These were defined as the best point-biserial correlations between the matrix element and group membership (control versus patients). Table 1a shows the characteristics selected.
2.3.2. Factor analysis
Factor analysis is a data reduction method which clusters sets of inter-correlated variables for a given population [31] to find a small number of independent higher order variables that describes the original set with minimal information loss [32]. As shown in Fig. 3, the sample covariance matrix was used in the principal component method (PCM) of factor analysis for the derivation of the factor loadings. The variables used in this analysis were the six characteristics on each of the orthogonal leads aVF, V2 and V5; hence 18 variables were used. Factors with an Eigen value greater than one were selected and rotated using the varimax transformation [31]. The three factors with the highest point-biserial correlation with group membership (control or patients) were selected. Table 1a shows these factors and the proportion of total variance accounted for by them was 35%. Table 1b shows the factor loadings. This three-factor
rotated solution then provided an input vector to the discriminant analysis on group membership:

\[ V_n = \{ \text{factor}_2, \text{factor}_6, \text{factor}_3 \} \] (6)

2.4. **Beat averaged P-waves: cardiological analysis**

Cardiological analysis refers to the typical measurements from the ECG trace that cardiologists use to examine the P-wave. The parameters chosen (P duration, terminal force, and P/PR ratio) are the parameters used to indicate the possibility of atrial abnormality. The P-wave duration is the time in milliseconds between the initial excitation of the sino-atrial node (onset) and the completion of depolarisation in the left atria (offset). Normal duration is considered to be less than 120 ms [7]. The P terminal force in V1 (PV1) is defined as the product of duration and amplitude of the P-wave's negative phase. The threshold of 0.04 mm s (4 mV ms) is commonly used as an indicator for the specific presence of LAE.

The (P/PR) ratio is the ratio of the P-wave duration over the PR segment duration. The PR segment duration is defined as the time from atrial excitation to the onset of ventricular excitation. A ratio of less than 1.0 has been shown to indicate right atrial enlargement. A ratio of greater than 1.6 has been shown to indicate left atrial enlargement. These limits reflect the alterations to the atrial conduction path caused by enlargement of either atrium. The P-wave duration for lead II, the P/PR duration for lead II, and the P-wave terminal force are extracted from the beat averaged P-waves for a single subject and formed into a single feature:

\[ V_n = \{ P, P_{PR}, PV1 \} \] (7)

2.5. **Linear discriminant analysis**

Fig. 4 illustrates the methodology for evaluating each of the four methods. Firstly the P-waves are extracted from the ECG for all normal and abnormal subjects and a single feature vector for each subject is generated. To examine the classification accuracy (sensitivity) of each of the four wavelet-based methods and the cardiological methods, linear discriminant analysis (LDA) was performed on each method separately. The issue of specificity (misclassifying a normal as abnormal) was not practical given the small sample of control subjects. Many studies only report classification accuracy for the data actu-

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Factor analysis</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>aVF median freq.</td>
<td>0.878</td>
<td>0.034</td>
</tr>
<tr>
<td>V2 median freq.</td>
<td>0.626</td>
<td>0.060</td>
</tr>
<tr>
<td>V5 median freq.</td>
<td>0.534</td>
<td>-0.278</td>
</tr>
<tr>
<td>aVF QV</td>
<td>0.015</td>
<td>0.913</td>
</tr>
<tr>
<td>aVF IQR</td>
<td>0.043</td>
<td>0.909</td>
</tr>
<tr>
<td>V2 peak value</td>
<td>-0.214</td>
<td>-0.022</td>
</tr>
<tr>
<td>aVF peak value</td>
<td>-0.214</td>
<td>-0.392</td>
</tr>
</tbody>
</table>

Bold values indicate the significant loadings.
ally provided to the classifier (i.e. training data). This results in an inflated estimate of how well the classifier will perform on data vectors that it has not been trained on. Therefore, a cross-validation set is required. The value in determining this generalizability is in the applied use of the discriminant equation, for example in a clinic where the P-wave analysis method is used on new cases. To estimate the sensitivity of each classifier the method as shown in Fig. 4 was used. The set of abnormal feature vectors were randomly separated into 2 sets (training and cross-validation) with each set containing 50% of the abnormal feature vectors. All of the normal controls were used in the “training” phase. In the training phase the LDA computes the optimum linear solution to maximize correct classifications [33]. The equation is then applied to the cross-validation set to determine the generalizability of the discriminant equation to “unseen” subjects from the

Fig. 5 – Mean and standard error of means (SEMs) of derived wavelet characteristics for normal and abnormal groups for the three-lead method. Characteristics are shown for leads aVF, V2 and V5 for (a) total energy, (b) peak value, (c) median frequency, (d) peak frequency, (e) inter-quartile range (IRQ) and quartile variation (QV). The discrimination capacity is a function of how many SEMs separate the means between the two groups.
abnormal group. The process of splitting the abnormal group randomly into two equal-sized subsets was repeated 100 times and the classification accuracy averaged over these 100 runs. The general form of the LDA equation is:

\[ D_n = \sum_{m=1}^{M} h_m V_{n,m} \quad 1 \leq n \leq N \]  

(8)

For the set of M characteristics used (indexed by m), the discriminant analysis derives a set of canonical coefficients, \( h_m \). The set of canonical coefficients, \( h \), of the linear discriminant equation were applied to the feature vectors of each case, \( n \), in the training set (set I) to produce a distance score, \( D_n \). The discriminating threshold, \( \theta \), is derived from the distance scores training set. The threshold is set at the point which maximally separates the distance scores of each group in the training set.

LDA is a good initial choice for a classifier because of its relative simplicity and good overall classifying generalization. The selection of a nonlinear discriminator is usually made after a linear discrimination has not performed well. In this case a linear discrimination was shown to perform reasonably well without the extra complexity of nonlinear (e.g. quadratic) functions. A well-known risk of using nonlinear discriminators, such as a neural network with hidden layers, is the tendency for the classifier to overfit on the training phase and can produce poor generalisation performance [22].

3. Results

Fig. 5 shows the mean and standard error of the mean (SEM) of the six parameters for the two groups on leads aVF, V2 and V5. When interpreting the differences between normal and abnormal groups this difference should be considered in terms of number of SEMs separation between the group means. For example, aVF peak value (Fig. 5b) shows several SEMs between the means, whereas total energy V5 (Fig. 5a) shows a little more than 2 SEMs. Hence aVF peak value is a better discriminator than total energy V5. These results help in understanding the appearance of the best classifier characteristics shown in Table 1a. The peak value (Fig. 5b) and the median frequency (Fig. 5c) of the energy distribution both showed excellent group differences and these appear in Table 1a as good individual discriminators. The variance-related characteristic, QV, showed some ability to discriminate the groups (Fig. 5f) but IQR (Fig. 5e) did not appear to have this ability.

The sign of the “r” coefficients in Table 1a indicates the direction of correlation between the characteristic and group membership. A positive correlation indicates that a higher value for any particular characteristic is more likely to be associated with the abnormal group. The correlations of median frequency in leads II and V1 are both negatively correlated with group membership. This result indicates that the majority of the P-wave’s energy is contained at higher frequencies in the normal group than it is in the abnormal group.

The LDA cross validation results are shown in Table 2 (individual P-wave approach) and Table 3 (signal-averaged P-wave approach). For the individual P-wave approach the cross-validation results showed greater than 80% correct classification for three of the wavelet methods with the two-lead method performing the best at 86.63%. The Frank XYZ model did not perform very well with a cross validation sensitivity of 68.63%. For the signal-averaged approach, the best was the XYZ model but it only achieved 68.56%. For the standard cardiological approach the best single measure was P-wave duration at 73.06% and all measures combined was 76.44%.

The individual P-wave approach classifications are shown in Fig. 6. Each graph shows the subjects across the X-axis with patients on the left represented as dots and normal controls on the right as open triangles. For patients their average cross-validation values are plotted and for normals (where the
Fig. 6 – Individual P-wave approach: individual subject classification are based on cross-validation for patients (dots) and training results for normals (triangles) for each of the four methods: (a) leads aVF V2 V5 (b) leads II V1, (c) factor analysis (d) XYZ analysis. A distance score from the horizontal decision line (at zero) is on the ordinate. Subjects are simply grouped together on the abscissa. Correct classifications are patients above and normals below the decision line.

sample size did not permit cross-validation) the test-set values are given for reference. Misclassifications for the abnormal group are below the discrimination threshold line (normalized to zero). A misclassification for a normal subject would be above this line.

Finally, the P-waves extracted for each subject do not show constancy in their D scores. Fig. 7 shows the results of LDA classification on each P-wave for each subject which can be seen as a column of symbols showing the spread in D scores. It was of interest to illustrate this for the set of all subjects using the two-lead model as it produced the best subject-median classification result. The variability in abnormal cases is particularly evident. As noted in Section 2.1, for many subjects there was a full 1 min of recording, hence these subjects have more data points. If a small sample of ECG is acquired and a patient has such variability in P-wave morphology (as here captured by its wavelet properties) this may not be representative. The importance of this is brought up in the discussion below.

4. Discussion

This study set out to determine whether characterisation of the P-wave with wavelet decomposition would result in useful classification of abnormality. The results clearly indicate that this has been achieved, particularly with the two-lead and three-lead subsets of the clinical 12-lead ECG configuration. In addition, the two-lead wavelet characterization showed a 10% improvement in classification rate over the typical cardiological parameters.

The results achieved using the signal-averaged P-waves were in general poorer than the individual P-wave characterization approach. This might be attributed to the signal average being more susceptible to variance in the P-wave morphology that "smears" during the averaging process losing some informative features. Fig. 7 provides some support for this hypothesis in that patient individual P-wave scores do indeed show considerable variation.
Looking at the wavelet-based two- and three-lead models in greater detail, it can be noted that the characteristics with the best discriminant potential are the median frequency and the peak frequency energy value. The median frequency is the centre of the wavelet scale ‘energy’ distribution. This is defined in \( Eq. (1) \) and relates the magnitude of the wavelet coefficients (expressed as ‘energy’) to the wavelet scale. \( Fig. 5c \) shows that normal P-waves have a higher frequency than the abnormal P-waves. This reflects the more compact shape of the normal P-wave. The duration measure used by cardiologists is sensitive to this morphological difference in abnormal P-waves as well. It is underscored by the good performance achieved by the duration measure on its own (nearly as good as the composite of all the cardiological measures). The conduction of the pulse through the atria takes longer (i.e. extended in duration) due to atrial enlargement. It should be noted, however, that there are individual differences in the orientation of the long-axis of the heart. This would introduce variance in the lead II estimate of duration. One way to examine this influence would be to include an echo cardiogram measure of long-axis orientation in tandem with the P-wave measures used here, and correlate them.

Only one measure of variance appeared in the best lead models and this was the QV measure on avF and lead II (\( Table 1a \)). It is also apparent, that for the three-lead model, in \( Fig. 5e \) and \( f \), the variance measures do not result in large group differences (only avF QV spans more than two SEMs). Although two of the variance measures did load on factor 3 of the factor analysis, this method did not perform quite as well as the models that did not include the IQR measure.

One wavelet method that did not perform as well as might be expected was the Frank lead XYZ method. The XYZ configuration did perform the best when the averaged P-wave was used as an input, though this performance was inferior to the single P-wave two- and three-lead wavelet methods and the classical ECG measures. Although the orthogonal set is similar to some extent to the avF, V2, V5 set, this is only approximate and indeed these derived leads may not have been completely orthogonal. A recorded Frank-lead trial may well show better performance than we have shown here. Finally, the best wavelet method predictor of all was the simplest one using only leads II and V1. Adding a third lead whether by a factor analytic method or a derived lead set does not appear to offer classification advantage and, in fact, may add noise. The two-lead set used II and V1. The lead II may be particularly important as it is sensitive to P-wave duration. \( Table 1a \) shows that four of the wavelet parameters were, in fact, based on this lead with median frequency capturing the duration information. The two-lead computational load for a modern PC is also quite practical, particularly once a sound discriminant equation is arrived at.

The present study used a modest sample size and has demonstrated the utility in using an automated P-wave classifier, showing it to outperform the usual duration and other cardiological measures by at least 10%. It is important however to extend the confidence in this analytic approach by conducting these analyses on a much larger data set. This can be a very costly enterprise if new data is collected, however online databases with ECG recordings are slowly becoming available [34]. While the effects of age on cardiological measures taken from normal subjects does not appear to change much beyond the mid 50s [21] it would inspire greater confidence to have a larger control sample – particularly for getting a measure of specificity. Clearly these findings, with a modest sample size, warrant further study to extend the generality to other laboratory and clinical settings. Furthermore, our suggestions of the relationship of P-wave anomalies to underlying atrial structure should be followed up in a replication ECG study that also includes echo cardiogram to more directly validate, on a case by case basis, possible structural abnormalities that correlate to wavelet-based P-wave morphology. This work is now underway in our laboratory in collaboration with a cardiology clinic.

To be useful to practicing cardiologists and GPs with ECG recording capabilities, the present system could be incorporated into current ECG software. This would be as a third-party plug in, for example to Cardioview\textsuperscript{TM}. A large set of cases should be compiled (future work on this is needed) to provide a robust LDA-trained and classified reference set. An interface could be easily added that allows the user to specify the patient, the database of LDA-trained cases and the form of the output, such as the distance scores for each of the diagnostics groups in the selected database as well as an interpretation of the P-wave morphology based on the wavelet parameter analysis.

In conclusion, the success of this wavelet-based approach may prove advantageous in the diagnosis of low amplitude and dispersed P-waves, which are commonly observed in elderly participants [11]. This can make diagnosis by visual inspection of the ECG a difficult task. Furthermore, the results of the wavelet methods described here compare favourably to the higher order system modelling approach [16] and are significantly better than the DFT method. One advantage of the present approach is that it does not rely on any initial assump-

---

**Fig. 7** – The two-lead model individual P-wave LDA results. Each individual P-wave from each subject was submitted to LDA and a line of best separation (middle horizontal line) computed. Each subject is apparent as a column of symbols (set of P-wave distance scores). Patients are to the left of the vertical line as x’s and normals to the right as diamonds.
tions about the generation of the P-wave and could therefore have wide applicability.

Conflict of interest

None.

Acknowledgements

The authors wish to acknowledge the support of Hearts 1st Cardiology Clinic in Greenslopes Private Hospital, for collecting and providing the ECG data, specifically, the assistance of Tony Forshaw, Geoff Spicer and Dr Roes Pascoe.

REFERENCES