Available online at www.sciencedirect.com
ScienceDirect
Journal of Electrocardiology 50 (2017) 781 – 786
JOURNAL OF Electrocardiology
www.jecgonline.com

A decision support system and rule-based algorithm to augment the human interpretation of the 12-lead electrocardiogram

Andrew W. Cairns, BSc, [a,*] Raymond R. Bond, PhD, [a] Dewar D. Finlay, PhD, [a] Daniel Guldenring, PhD, [a] Fabio Badilini, PhD, [b] Guido Libretti, BSc, [b] Aaron J. Peace, PhD, [c] Stephen J. Leslie, PhD, FRCP [d]

[a] *Ulster University, Northern Ireland, UK*
[b] *AMPS-LLC, New York, USA*
[c] *Altnagelvin Hospital, Northern Ireland, UK*
[d] *NHS Highland, UK*

Abstract

Background: The 12-lead Electrocardiogram (ECG) has been used to detect cardiac abnormalities in the same format for more than 70 years. However, due to the complex nature of 12-lead ECG interpretation, there is a significant cognitive workload required from the interpreter. This complexity in ECG interpretation often leads to errors in diagnosis and subsequent treatment. We have previously reported on the development of an ECG interpretation support system designed to augment the human interpretation process. This computerised decision support system has been named 'Interactive Progressive based Interpretation' (IPI). In this study, a decision support algorithm was built into the IPI system to suggest potential diagnoses based on the interpreter's annotations of the 12-lead ECG. We hypothesise semi-automatic interpretation using a digital assistant can be an optimal man-machine model for ECG interpretation.
Objectives: To improve interpretation accuracy and reduce missed co-abnormalities.
Methods: The Differential Diagnoses Algorithm (DDA) was developed using web technologies where diagnostic ECG criteria are defined in an open storage format, Javascript Object Notation (JSON), which is queried using a rule-based reasoning algorithm to suggest diagnoses. To test our hypothesis, a counterbalanced trial was designed where subjects interpreted ECGs using the conventional approach and using the IPI + DDA approach.
Results: A total of 375 interpretations were collected. The IPI + DDA approach was shown to improve diagnostic accuracy by 8.7% (although not statistically significant, p-value = 0.1852), the IPI + DDA suggested the correct interpretation more often than the human interpreter in 7/10 cases (varying statistical significance). Human interpretation accuracy increased to 70% when seven suggestions were generated.
Conclusion: Although results were not found to be statistically significant, we found; 1) our decision support tool increased the number of correct interpretations, 2) the DDA algorithm suggested the correct interpretation more often than humans, and 3) as many as 7 computerised diagnostic suggestions augmented human decision making in ECG interpretation. Statistical significance may be achieved by expanding sample size.

Keywords: Decision support; 12-lead Electrocardiogram; Interpretation; Algorithm; Diagnoses; ECG criteria

Introduction

Cardiac abnormalities are often manifested in a 12-lead Electrocardiogram (ECG) [1]. However, due to the complex nature of 12-lead ECG interpretation including analysis of multifarious leads, deflections and patterns, a significant cognitive workload is required from the interpreter [2]. This is in addition to the interpreter having to refer to an intricate knowledge-base in cardiac pathology and cognitively cross-referencing a large set of ECG criteria. This complexity in ECG interpretation often leads to errors in diagnoses and treatment. Diagnostic accuracy has been reported to be as low as 40% [3–5]. It has been shown that even expert clinicians can act impulsively in providing a diagnosis based on their first impression [5–9] and, in turn, potentially miss co-abnormalities due to 'early satisfaction syndrome' and

* Corresponding author. Room 16J27, University of Ulster, Jordanstown Campus, Shore Road, Newtownabbey, Co., Antrim BT37 0QB, UK.
E-mail address: cairns-a3@email.ulster.ac.uk

http://dx.doi.org/10.1016/j.jelectrocard.2017.08.007
0022-0736/

cognitive biases such as anchoring or confirmation bias [8–10]. It is therefore imperative to optimise how physicians interpret the 12-lead ECG.

The 12-lead ECG is often presented alongside a computerised diagnosis to assist the interpreter. However, this computerised analysis is firstly, often inaccurate, as machine algorithms often fail to recognise patterns in noisy ECG signals [11–14]. Secondly, with common computerised analysis often being inaccurate [12,14,15] and only providing a single proposed diagnosis, cognitive biases can be incurred, including; 1) anchoring bias (fixation on a premature suggestion), 2) confirmation bias (interpreters seek diagnostic features to confirm, rather than falsify, a diagnosis), or 3) premature closure (an interpreter accepting a diagnosis before verification) [15,16]. Therefore, numerous studies have recommended computerised ECG interpretation should always be accompanied by clinical human affirmation [17,18].

Whilst the 12-lead ECG presentation has remained unchanged for more than 70 years, [19] medical practices, including the NHS in the UK, are striving towards a paperless environment [20]. This provides an opportunity to use interactive computing and touch screens to aid the human interpretation of the ECG. Opportunities such as this, as well as the motivation to integrate the human interpreter in the decision making process, have inspired our previous work where we developed an approach that provided a set of interactive questions and prompts to guide an interpreter through the ECG reporting process. The model was named 'Interactive Progressive based ECG Interpretation' (IPI) [21]. Coupling this previous research with motivations to further increase diagnostic accuracy, and reduce cogitative bias, we have augmented the IPI model with a rule-based algorithm to generate multiple computerised diagnostic proposals.

Thus, we hypothesise that semi-automatic interpretation is an optimal man-machine model for ECG interpretation. This hypothesis is based on the fact that the human cognitive memory prevails in pattern recognition (i.e. in noisy signals) enabling the interpreter to provide more accurate annotations whilst a machine performs better at using annotations to reason against a large set of rules (ECG criteria).

Methodology

A differential diagnosis algorithm (DDA) has been integrated into the IPI system to provide multiple potential ECG diagnoses based on a human interpreter's ECG annotations (feature detection, waveform measurements and segment analysis). The number of suggestions generated by the DDA varies depending on human annotations. To accomplish this, the rule-based algorithm evaluates an interpreters' response to each inputted annotation and matches these annotations against a set of ECG diagnostic criteria. As a de-biasing strategy, multiple potential ECG diagnoses are presented following the interpreter inputting their annotations. The algorithm will only provide a diagnostic suggestion when the annotations match at least 50% of the diagnostic criteria. Pseudo code for the DDA can be seen in Fig. 1, and a screenshot of the system can be seen in Fig. 2. The model and algorithm design is further described in [22].

The algorithm was implemented using web technologies including JavaScript, PHP, HTML and CSS. To store diagnostic criteria, the system uses the device agnostic data model and storage format known as JavaScript Object Notation (JSON) for defining the rules. These rules are then queried using the decision support algorithm programmed in JavaScript.

Study design

A counterbalanced study design was used to compare the diagnostic accuracies achieved when interpreters use the IPI + DDA system in comparison to the conventional approach to reading ECGs (i.e. all 12 leads presented in the commonly accepted 3 × 4 + 1R format). Therein, each interpreter interpreted five ECGs using the conventional method and five ECGs using the IPI method. The entire cohort was split into two subgroups referred to as A and B. Group A interpreted ECG numbers 1–5 using the conventional method and ECGs numbered 6–10 using the IPI + DDA method. Conversely, group B interpreted ECGs 1–5 using the IPI + DDA method, and ECG 6–10 using the conventional method. All interpreters were asked to provide a self-assessed confidence rating for each interpretation (scale 1–10, where 10 = very confident).

All chosen ECGs (ten) derived from a publically available ECG repository with predefined pathologies and interpretation difficulty rankings [23] and were selected to align with the UKs National Health Service (NHS) healthcare science practitioner training programme [24] and to express the European Society of Cardiology (ESC) Core Curriculum for the General Cardiologist [25]. Seven of the ten ECGs exhibit cardiac pathologies (e.g. STEMI) whilst the remaining three ECGs exhibit anomalies (e.g. lead misplacement or dextrocardia).

Recruitment

Recruitment resulted from convenience sampling from four available participant cohorts; 1) International Society for Computerized Electrocardiology (ISCE) delegates, 2) junior doctors in two Scottish NHS trusts, 3) clinical physiology students and 4) European Society of Cardiology members (ESC). Participation was undertaken in both a classroom environment and remotely via website hyperlinks. As the system was developed responsively using web technologies, it is device and platform agnostic and can be accessed with any device with an internet connection.

Data collection

Before beginning, interpreter demographics were collected using an online form. These include; age, gender, occupation, years of experience interpreting ECGs and number of ECGs interpreted annually. Interpreters were also required to give informed consent before proceeding to interpret all 10 ECGs. All annotations are collected and saved via an AJAX function

```
Algorithm 1: Suggestion algorithm
Input: Input click event for each question
begin
  foreach input event do
    Empty suggestion element placeholder
    Data: NewVariable ⟵ Value         /* New variables based on current data values */
    Data: NewVariableArrays ⟵ []      /* Empty variable arrays (populated later) */
    Data: Call to ECG JSON data                            /* .getJSON function */
    foreach variable do
      if variable = ECG criteria (crit) then
        RelevantECGarray ⟵ crit
    foreach ECGarray do
      MasterArray ⟵ ECGarray
    Data: Create ⟶ MasterObject

    Array to object conversion loop to enable a frequency count
    foreach ECGarray in MasterArray do
      foreach ECGdiagnoses in ECGarray do
        if ECG diagnoses == false (i.e. does not exist / empty / or 0) then
          MasterObject ⟵ 0
        else if Diagnosis is unique then
          Add diagnoses to MasterObject and add 1 to the count
          MasterObject ⟵' Diagnosis' : Count = 1
        else
          Add count of 1 to relevant Diagnosis
          MasterObject ⟵' RelevantDiagnosis' : Count + 1
    Convert MasterObject ⟶ SortedArray
    Sort (decending) SortedArray
    foreach Diagnois in SortedArray do
      Output: Diagnosis
```

Fig. 1. Pseudo code illustrating the DDA which was implemented using JavaScript. The code illustrates data collection, a number of data–specific criteria searches, loops to generate suggestion list and final suggestion display.

in a MySQL database on an Apache web server. In total, there were 49 interpreters included in this study. Of which, 35 participant completed ECG interpretations using both approaches, whilst 14 participants did not complete interpretations using both approaches but their completed interpretations were recorded. This resulted in 280 interpretations from 35 participants (as some participants did not complete all ECGs), plus 70 interpretations from 14 participants who did not use both methods. Overall 375 interpretations were recorded (215 control interpretations, 160 IPI + DDA interpretations).

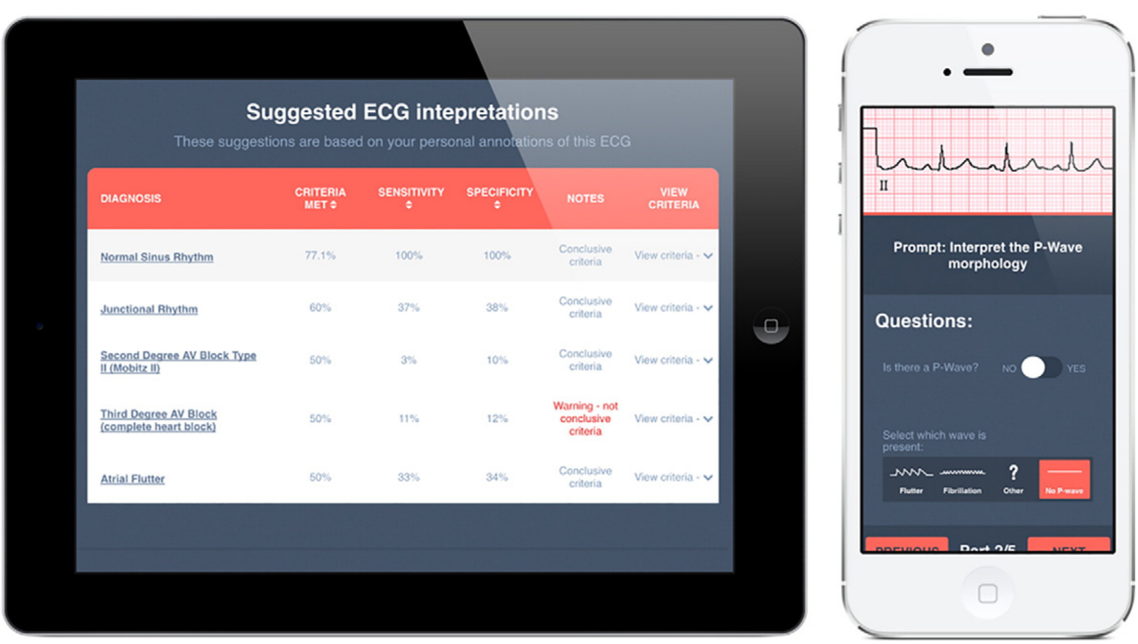

Fig. 2. Presentation of the IPI+DDA system on mobile devices. An example of generated suggestion displays, questions and prompts.

Table 1
Table illustrating the number of diagnostic suggestions from the IPI+DDA, number of correct algorithm diagnoses, number of instances (i.e. the number of times the relative number of suggestions was generated), and the percentage of instances containing the correct suggestion.

Number of diagnostic suggestions from the IPI + DDA	Number of instances	Number of instances that contain a correct diagnostic suggestion	Percentage of correct algorithm diagnosis (%)
1	6	0	0
2	8	5	62.5
3	15	4	26.7
4	21	3	14.3
5	18	5	27.8
6	14	4	28.6
7	10	5	50
8	8	5	62.5
9	12	9	75
10	8	1	12.5
11	11	6	54.6
12	8	6	75

Table 2
Table illustrating the number of suggestions compared to the number of correct human interpretations, the number of suggestion instances (i.e. the number of times the relative number of suggestions was generated), and the relative human accuracy as a percentage.

Number of diagnostic suggestions from the IPI + DDA	Number of instances	Number of instances that contain a correct human interpretation suggestion	Percentage of human accuracy (%)
1	6	1	16.7
2	8	4	50
3	15	7	46.7
4	21	6	28.6
5	18	9	50
6	14	9	64.3
7	10	7	70
8	8	2	25
9	12	4	33.3
10	8	1	12.5
11	11	5	45.5
12	8	3	37.5

Statistical analyses

Data was analyzed using the Structured Query Language, Microsoft Excel [26] and the R programming language [27]. A Shapiro–Wilk test was used to test for normal distribution, data was found to be not normally distributed. A two-tailed Mann-Whitney U test was used to test for statistical significance between interpretation methods. To compare statistical significance between interpretation method proportions we conducted Chi-squared tests. The p-value used to determine statistical significance was ≤ 0.05.

Results

The percentage of correct interpretations for reading ECGs using the conventional approach was 42.61% whilst interpretations using the IPI + DDA method was 51.35% (chi-squared p-value = 0.1852). Thus, interpretations resulting from use of the IPI + DDA were 8.7% more accurate.

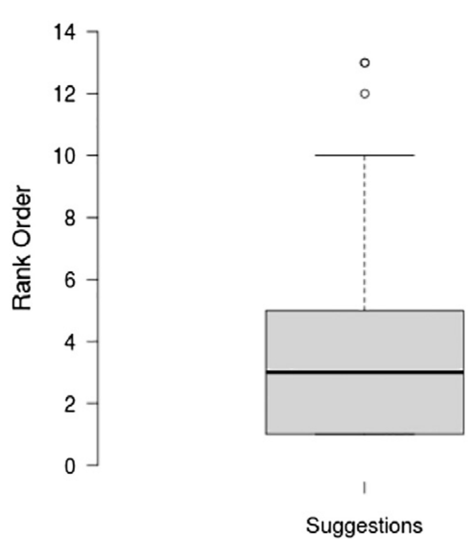

Fig. 3. Box plot illustrating the range and median of the correct diagnosis rank order in the list of suggestions generated by the DDA.

The IPI method did not improve the detection of ECGs which had been recorded where there was lead misplacement or dextrocardia despite the IPI + DDA interface directly prompting users to carry out an inspection for lead misplacement. Overall self-rated confidence in ECG interpretation using the control method was 5.37/10 (SD = 2.95) whilst the IPI method was 5.58/10 (SD = 3.02). This indicates interpreters feel 3.9% (although not statistically significant, Z-Score = −0.7, p-value = 0.48) more confident in interpreting ECGs using the IPI method. The average duration of interpretations using the conventional method was 108.55 s (SD = 32.57) and 629.94 s (SD = 266.98) when using the IPI method. Thus, the average IPI method duration was 6.19 times longer. However, the 6-fold increase in interpretation time is confounded by participants being unfamiliar users of the system. This is highlighted in further interpretation time analysis. As users become more familiar with the system interpretation time decreases (mean time shortening = 130.25 s).

Correct suggestion ranks of the decision support algorithm

Due to the DDA design, there is a variable number of suggestions listed based on interpreter input. However, we found that between 3 and 6 suggestions were most frequently presented (44% of all interpretations). The mode rank of the correct suggestion in the list was 3 (mean = 3.63, SD =3.01). The correct suggestion appeared within the first three suggestions in 60% of interpretations (refer to Table 1 and Fig. 3).

Algorithm accuracy vs. number of suggestions

We found that when 2 suggestions are presented, there is a 62.5% likelihood the right suggestion will be in the list. However, we also found that when 9 suggestions are generated there is a 75% likelihood of the correct suggestion appearing in the list.

Table 3
Table illustrating the percentage difference in accuracy between the IPI+DDA method and the human interpreter in ECG interpretation. Positive inflection illustrates the algorithm is more accurate, conversely a negative inflection illustrates human interpretation was more accurate.

ECG number	Percentage difference in accuracy between DDA and the human interpreter (positive = algorithm more accurate, negative = human more accurate)	Test of equal or given proportions (chi-squared)
STEMI	10.53	p-Value = 0.7271
LVH	−37.50	p-Value = 0.06789
RAE	31.25	p-Value = 0.1365
VT	−28.57	p-Value = 0.2519
SVT	−50.00	p-Value = 0.009598
Atrial fibrillation	5.88	p-Value = 1
Limb lead misplacement	22.22	p-Value = 0.2291
Dextrocardia	25.00	p-Value = 0.1742
Chest lead misplacement	26.67	p-Value = 0.1709
Normal sinus rhythm	28.57	p-Value = 0.1052

Human accuracy vs. number of suggestions

When comparing human interpretation accuracy with varying number of suggestions generated by the DDA, we found that the human interpreter will provide the correct interpretation 70% of the time when seven suggestions are generated. When two, three, five or six suggestions are generated the human interpreter is more than 45% likely to interpret the ECG with a correct answer. This percentage is greater than the percentage of correct human interpretations when using the conventional method of ECG interpretation. More details are reported in Table 2.

Algorithm accuracy vs. human accuracy

When comparing algorithm suggestions directly with the human interpretations for each ECG we find in 7/10 cases the DDA algorithm provided more correct interpretations than the human interpreter (varying statistical significance, refer to Table 3). However, human interpretation was more accurate when reading ECGs exhibiting Left Ventricular Hypertrophy (LVH), Ventricular Tachycardia (VT) and Supraventricular Tachycardia (SVT). In the case of LVH, one possible reason for this is that the system does not require input for QRS amplitude. Therefore, the criteria for LVH is incomplete resulting in the algorithm being unable to process relevant data to generate an accurate suggestion. Similar, assumptions can be made in the cases of VT and SVT.

Discussion

There is potential to improve the accuracy of ECG interpretation by using an interactive decision support system to augment the human interpretation process. We found the IPI + DDA system increased the number of correct interpretations by 8.7% and improved interpreter self-rated interpretation confidence by 3.9% (although results were not statistically significant). In 70% of cases the IPI + DDA suggested the correct interpretation more often than the human interpreter. With the ability to augment the interpretation process with potential diagnoses, we identified that displaying as many as seven computerised diagnoses improves human diagnostic accuracy in ECG interpretation.

The IPI + DDA system was not compared against results from a previous study in which the IPI model was used without the DDA [21], for a number of reasons; participants were from different cohorts, each study had a different experimental design (two arm vs one arm counter balance), the starting ability is superior in IPI cohort. With this all in mind, we have noted that overall accuracy did not improve between the IPI and IPI + DDA methods.

Limitations include relatively small numbers of ECGs used within the system for this study. Also, a relatively small number of interpreters with varied experience was also present. However, a respectable number of ECG interpretations was recorded. As a result, the statistical comparisons are widely not significant, which weakens any definitive conclusions. A further limitation is the lack of control in the gold standard for the ECG diagnoses used with the study.

Numerous adaptations could be made to enhance this system. Refinements could be made to the diagnostic criteria stored in the JSON object, for example, adding further specific criteria to help diagnose LVH, VT and SVT. A second enhancement could be to define and implement weightings to correspond with the importance of each diagnostic criterion in the JSON object allowing the DDA algorithm to improve how it rank its suggestions. Thirdly, some annotations could be pre-calculated by accurate computerised analysis, this could decrease interpretation time and increase diagnostic accuracy. One further enhancement could be to create an interface to allow clinicians to edit/update diagnostic criteria following a verification process.

Conclusion

Although results were not statistically significant, we found; 1) our decision support system increased the number of correct interpretations, 2) the DDA algorithm suggested the correct interpretation more often than humans, and 3) as many as 7 computerised diagnostic suggestions augmented human decision making in ECG interpretation. Statistical significance may be reached by expanding the sample size. With future of ECG interpretation likely to be paperless, there is an opportunity to improve ECG interpretation accuracy using an interactive decision support system.

Acknowledgements

This research was funded by the Department of Employment and Learning (DEL).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jelectrocard.2017.08.007.

References

[1] World Health Organisation. "WHO|cardiovascular diseases." [online]. Available: http://www.who.int/topics/cardiovascular_diseases/en/.

[2] Sibbald M, de Bruin ABH, van Merrienboer JJG. Checklists improve experts' diagnostic decisions. Med Educ 2013;47(3):301–8.

[3] Peace A, et al. Using computerised interactive response technology to assess electrocardiographers and for aggregating diagnoses. J Electrocardiol 2015;48(6):995–9.

[4] Novotny T, et al. Data analysis of diagnostic accuracies in 12-lead electrocardiogram interpretation by junior medical fellows. J Electrocardiol 2015;48(6):988–94.

[5] Novotny T, et al. The role of computerized diagnostic proposals in the interpretation of the 12-lead electrocardiogram by cardiology and non-cardiology fellows. Med Inform 2017;101:85–92.

[6] Novotny T, et al. Data analysis of diagnostic accuracies in 12-lead electrocardiogram interpretation by junior medical fellows. J Electrocardiol 2015;48(6).

[7] Bond RR, et al. Eye tracking in the assessment of electrocardiogram interpretation techniques. 2012 Computing in cardiology; 2012. p. 581–4.

[8] Breen CJ, Bond R, Finlay D. An evaluation of eye tracking technology in the assessment of 12 lead electrocardiography interpretation. J Electrocardiol 2014;47(6):922–9.

[9] Wood G, Batt J, Appelboam A, Harris A, Wilson MR. Exploring the impact of expertise, clinical history, and visual search on electrocardiogram interpretation. Med Decis Making 2014;34(1):75–83.

[10] Bond RR, et al. Assessing computerized eye tracking technology for gaining insight into expert interpretation of the 12-lead electrocardiogram: an objective quantitative approach. J Electrocardiol 2014;47(6):895–906.

[11] Bogun F, et al. Misdiagnosis of atrial fibrillation and its clinical consequences. Med 2004;117(9):636–42.

[12] Shah AP, Rubin SA. Errors in the computerized electrocardiogram interpretation of cardiac rhythm. J Electrocardiol 2007;40(5):385–90.

[13] Willems JL, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. Med 1991;325(25):1767–73.

[14] Estes NAM. Computerized interpretation of ECGs: supplement not a substitute. CircArrhythm Electrophysiol 2013;6(1):2–4.

[15] Tsai TL, Fridsma DB, Gatti G. Computer decision support as a source of interpretation error: the case of electrocardiograms. J Am Med Inform Assoc 2003;10(5):478–83.

[16] Cruz MF, Edwards J, Dinh MM, Barnes EH. The effect of clinical history on accuracy of electrocardiograph interpretation among doctors working in emergency departments. Aust 2012;197(3):161–5.

[17] Goodacre S, Webster A, Morris F. Do computer generated ECG reports improve interpretation by accident and emergency senior house officers? Postgrad Med J 2001;77(909):455–7.

[18] Kligfield P, et al. Recommendations for the standardization and interpretation of the electrocardiogram: part II: electrocardiography diagnostic statement list a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council. J Am Coll Cardiol 2007;49(10):1128–35.

[19] Madias JE. The 13th multiuse ECG lead: shouldn't we use it more often, and on the same hard copy or computer screen, as the other 12 leads? J Electrocardiol 2004;37(4):285–7.

[20] TechUK. Digitising the NHS by 2018 — one year on. [Online]. Available: http://jac.co.uk/wp-content/uploads/2013/03/Digitising_the_NHS_-_One_Year_On.pdf 2014.

[21] Cairns AW, et al. A computer-human interaction model to improve the diagnostic accuracy and clinical decision-making during 12-lead electrocardiogram interpretation. J Biomed Inform 2016;64:93–107.

[22] Cairns Andrew W, Bond Raymond R, Finlay Dewar D. An annotation driven rule-based algorithm for suggesting multiple 12-lead ECG interpretations; 2016.

[23] Casella L, Mangat I, Nader A. ECG made simple. [Online]. Available: http://www.ecgmadesimple.com/ 2015.

[24] NHS. Practitioner training programme; 2014.

[25] Gillebert TC, et al. ESC core curriculum for the general cardiologist (2013). Eur Heart J 2013;34(30):2381–411.

[26] Microsoft. Microsoft Excel|spreadsheet software. [Online]. Available: https://products.office.com/en-gb/excel.

[27] RStudio. RStudio|open source and enterprise-ready professional software for R. [Online]. Available: https://www.rstudio.com.