MEDINFO 2001
V. Patel et al. (Eds)
Amsterdam: IOS Press

Employment of Intra-Individual Variability to Improve Computerized ECG Interpretation

Bob J.A. Schijvenaars, Jan A. Kors, Gerard van Herpen, Jan H. van Bemmel

Department of Medical Informatics, Erasmus University Rotterdam, The Netherlands

Abstract

One of the reasons for the limited practical utility of computer programs for interpretation of electrocardiograms (ECGs) is their susceptibility to intra-individual variability. Two of the most prominent sources of intra-individual variability in ECGs, electrode placement variations and respiration, were studied for their effects on computerized ECG interpretation. Previous research has shown that the effects of intra-individual variability on computerized ECG interpretation depend largely on the individual ECG. To enable the assessment of chest electrode position variations for individual standard 12-lead ECGs, ECGs resulting from simulations of such position variations were interpreted. Variability due to respiration was assessed by interpretating all individual ECG beats instead of an averaged beat.

In this paper two methods are presented that employ information about the intra-individual variability in individual ECGs. The first method provides an estimate of the reliability of the interpretation, the second attempts to improve the interpretation itself.

In the first method we quantified the variation in interpretation caused by the two sources of intra-individual variability with the use of a stability index, a high index value indicating a low variation in interpretation. This index was subsequently studied using two sets of ECGs. For the first set a 'clinical' reference interpretation was obtained from discharge letters. For the second set three cardiologists provided a 'cardiologists' reference. The performance of subgroups of ECGs having stability indices higher than a particular value was computed. It appeared that for the 'cardiologists' reference, the interpretations of ECGs with a high stability index were more often correct. No effect was found for the 'clinical' reference.

In the second method we attempted to improve the original interpretation by combining the alternative interpretations into a new interpretation. This was done by taking the median or the average of the quantified alternatives. These combined interpretations proved to perform better than the original interpretation when a cardiologist's interpretation was taken as a reference.

This paper shows that intra-individual ECG variability can be used to improve original interpretations. This can be done without having to record multiple ECGs, provided that a model is available to simulate intra-individual variability. The presented methods do not depend on the classification algorithm that is used. They can be used both during classifier design to correct imperfections, and in routine use of the classifier to produce more representative classifications.

Keywords:

Computer-Assisted Diagnosis; Electrocardiography; Body Surface Potential Mapping

Introduction

Despite their good diagnostic performance, computer programs for the interpretation of electrocardiograms (ECGs) suffer from a number of drawbacks that limit their practical utility [1,2], one of the most important being the vulnerability for individual ECG variability [1,3]. Identical ECG signals will result in identical measurements and interpretations, but small (and diagnostic inconsequential) differences between signals may result in an entirely different diagnostic interpretation [3,4].

One of the reasons that this vulnerability still exists is that ECG computer programs are often only validated with respect to their accuracy (e.g., sensitivity and specificity). However, the stability of the interpretation is a key factor for user acceptance as well. Especially in serial ECG interpretation where multiple interpretations are performed on different ECGs from the same individual, a large variation in interpretations will reduce the confidence of the user. This interpretation variability is often caused by intra-individual ECG variability.

Various sources of variability may affect the shape of the electrocardiogram (ECG). Non-pathological, 'circumstantial' variability can be related to technical sources (e.g., electrode positioning, mains interference or patient posture) or to biological sources (non-cardiac muscle activity, respiration). In the past, attention has been paid to assessing [5] and minimizing [6,7] ECG variability resulting from these sources.

A literature review [8] revealed that two of the most important sources of intra-individual ECG variability encountered in daily practice are variations in electrode positioning and respiration. The latter results in variations within a single ECG, the former in variations between ECGs. We assessed the effects of chest electrode position variations using Body Surface Potential Maps (BSPMs). The effects on measurements and diagnostic interpretations were shown to highly vary from ECG to ECG [9]. Assessment of the effects of intra-individual ECG variability, therefore, must be done for each individual ECG separately.

The effects of respiration on individual ECGs can be easily determined by assessing the variability between beats in the same ECG (beat-to-beat variability). Assessing the effects of chest electrode position changes based on a single 12-lead ECG is not straightforward, however. In a previous study [10], we determined transformations that approximated ECGs recorded from mispositioned chest electrodes using ECGs from correct electrode positions. These transformation matrices proved to produce reasonably accurate estimations of the effects of chest electrode displacement.

Using the above methods, intra-individual ECG variability can now be quantified. This paper presents two methods to employ this information: 1) to estimate the reliability of the interpretation of an individual ECG, and 2) to improve this interpretation.

Materials and Methods

Given an ECG, alternative ECGs can be generated for each of the two sources of variability, electrode position variation and respiration. In case of beat-to-beat variability, each individual beat is considered an alternative ECG. For variability caused by electrode position variation alternative ECGs were generated with the previously developed transformation techniques [10]. Figure 1 shows the configuration used for these electrode displacement simulations. Five downward and five upward displacements of all six chest electrodes were simulated, resulting in 10 simulated ECGs.

Both the original and all alternative ECGs were processed by our Modular ECG Analysis System (MEANS). MEANS has extensively been tested, both by the developers themselves [11] and by others [12]. For each diagnostic category, MEANS assigned one of four qualifiers: 'absent', 'possible', 'probable' and 'definite', coded as 0, 1, 2, or 3, respectively. For this study, only the diagnostic category myocardial infarction (MI) was considered.

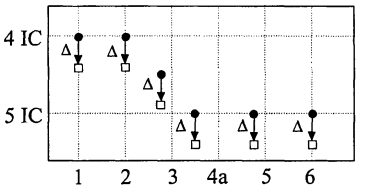

Figure 1 - The displacements used in the simulation of the chest electrode position shifts. The row labels indicate the fourth and fifth intercostal (IC) space. The black dots indicate the standard precordial electrode positions; the open squares the positions after displacement Δ.

Thus, for each individual ECG, two sets of qualifier codes are obtained, one set consisting of the codes of the individual beats of the original ECG, the other containing the codes of the simulated displaced ECGs.

Interpretation reliability

The variability in a set of n coded interpretations was quantified using a stability index I, defined as:

$$I = 1 - \frac{1}{3n}\sum_{i=1}^{n}|x_i - x_0| \qquad (1)$$

with x_0 the qualifier code of the original interpretation, x_i the code of interpretation i, and $3n$ a normalization factor to make the index independent of the number of observations (e.g., the number of beats in a recording may vary). Maximum stability (i.e., all equal codes) yields a stability index one, minimum stability (the original code is an outlier) an index zero.

We hypothesized that the stability index correlates with the performance, i.e., sensitivity and specificity, of the classification algorithm, a high index value indicating a more reliable interpretation than a lower value. To test this hypothesis, we used two sets of ECGs, as will be described below.

Interpretation improvement

Two methods were used to derive a new interpretation from a set of alternative interpretations: taking the median and taking the average. Both these methods were applied to the electrode position set of alternative interpretations and to the beat-to-beat set of interpretations. This resulted in four combined interpretations (EP-median, EP-average, BB-median, and BB-average). A fifth interpretation was computed by taking the median of the original interpretation and the two median interpretations (Fin-median). A sixth combined interpretation was computed by taking the average of the original and the two combined averaged interpretations (Fin-average). The performance of all six combined interpretations was compared with that of the original interpretation to test whether the combined interpretations showed improved performance.

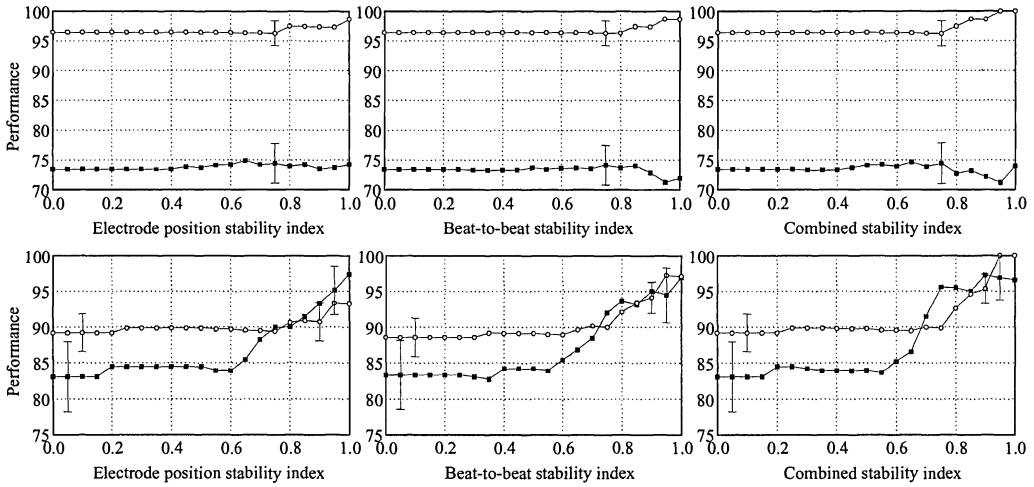

Figure 2 – Classification performance on subgroups of ECGs from the 'clinical' database (top row) and the 'cardiologists' database (bottom row) with stability indices larger than a particular value. Performance is expressed by sensitivity (solid squares) and specificity (open circles) and is plotted for the electrode position stability index (leftmost column), the beat-to-beat index (middle column) and the combined stability index (rightmost column). The vertical bars denote standard errors.

Material

To test whether the stability index is correlated with the performance of the original interpretation, two sets of ECGs were used. A first set of 272 12-lead ECGs (mean age 57 years; 35% female) was selected from a large database of ECGs that had routinely been recorded in the cardiology department of the university hospital in Rotterdam, The Netherlands. A reference interpretation for this set was determined by a cardiologist, based on the discharge letter. Prevalences according to this reference interpretation were 69.1% for myocardial infarction (MI) and 30.9% for non-MI (normal and other abnormalities). This set will be referred to as the 'clinical' database. For the reference interpretation of the ECGs in this database only qualifier codes 0 ('absent') and 3 ('definite') were used.

A second set of 198 ECGs (mean age 56 years; 39% female) was collected during clinical routine in cardiology departments in five European hospitals and was interpreted by three cardiologists. Their interpretations were quantified in the same way as done for the computer-generated interpretations: using qualifier values 0, 1, 2 and 3. The average kappa value of the cardiologists' interpretations was 0.72, indicating substantial agreement. A reference classification was computed by taking the rounded average of the three qualifier codes for each ECG. This set is referred to as the 'cardiologists' database. According to the reference interpretation, prevalences were 27.3% for MI, and 72.7% normal or other abnormalities.

The hypothesis that a combined interpretation performs better than the original interpretation was tested using a database of 77,169 ECGs recorded in the cardiology department of the university hospital Rotterdam, and 16,841 ECGs recorded in a cohort study among apparently healthy individuals [13].

Results

Interpretation reliability

Sensitivity and specificity were used as measures of interpretation performance. Qualifier codes 2 ('probable') and 3 ('definite') were regarded as positive interpretations, codes 0 and 1 as negative ones. Figure 2 shows sensitivity and specificity for ECGs in the 'clinical' (top row) and 'cardiologists' (bottom row) databases that have a stability index larger than or equal to the values indicated on the abscissa. Thus, performance estimates at stability index 0 apply to all ECGs in the set, and estimates at index 1 apply to the most stable ECGs only. For the 'clinical' database a slight, but not important, increase in specificity can be observed for high stability indices though the standard errors are rather large. For the 'cardiologists' database the effect is markedly different. Both sensitivity and specificity increase considerably with increasing stability index.

Interpretation improvement

Only the ECGs where at least one of the combined interpretations differed from the original one were selected for further analysis since only these ECGs may cause a

performance difference. Table 1 shows the number of ECGs where a difference was observed.

Table 1. Number (and percentage) of ECGs with a combined interpretation that differed ±1, ±2, or ±3 qualifier points from the original interpretation.

Interpretation	±1	±2	±3
BB-median	2,135 (2.3)	976 (1.0)	1,064 (1.1)
EP-median	1,714 (1.8)	623 (0.7)	700 (0.7)
Fin-median	811 (0.9)	302 (0.3)	319 (0.3)
BB-average	8,404 (8.9)	1,277 (1.4)	131 (0.1)
EP-average	7,421 (7.9)	1,071 (1.1)	79 (0.1)
Fin-average	4,347 (4.6)	103 (0.1)	0 (0.0)

In 3,170 of the 94,010 ECGs, at least one of the combined interpretations showed a large difference (i.e., a qualifier code difference of 2 or 3) with the original interpretation. From this set, 182 ECGs were randomly selected and interpreted by an experienced cardiologist: 143 cases were classified as non-MI and 39 as MI. Taking his interpretation as the reference, the number of better, equal, and worse combined interpretations as compared to the original interpretation was determined (Table 2).

Table 2. The number of cases in which the interpretation improved, remained the same, or deteriorated compared to the original interpretation.

Interpretation	better	equal	worse
Original	0	182	0
BB-median	126	0	56
EP-median	115	18	49
Fin-median	115	18	49
BB-average	119	15	48
EP-average	103	31	48
Fin-average	46	120	16

Discussion

Our results demonstrate that both sensitivity and specificity increase considerably in subgroups of ECGs with a maximum stability index, i.e., a minimum intra-individual variability. In addition, all combined interpretations yield at least twice as many improvements as deteriorations.

These results show that the stability index can be used as a measure for the reliability of the original interpretation: ECGs with a high stability index are more often interpreted correctly. A possible explanation for this phenomenon is that the interpretation algorithms of MEANS make use of complex decision trees that subdivide and label a high-dimensional feature space. Ideally, the trees should present gradual transitions between decision regions, but due to the high dimensionality and the fact that trees use discrete thresholds, this is probably not accomplished for all regions. An insignificant change in one measurement might thus cause a move from one category to another. ECGs with a low stability index are probably located near decision region edges. This makes it more likely that the original classification is located in the wrong decision region.

The sensitivity on the 'clinical' database did not improve for higher stability indices. A possible reason might be the inherent difference between a diagnosis based on only the ECG and one using non-ECG clinical information. This phenomenon was also observed in the CSE study, where two different types of reference were used, but now on the same set of ECGs [12]. The consensus interpretation of a panel of cardiologists had a total accuracy of 79.2% when taking an ECG independent interpretation as a reference. Since the MEANS program was developed to mimic a cardiologist's interpretation, improvement is more likely to be seen with a similar reference than with an ECG-independent reference.

The use of decision tree algorithms may also be the explanation of the better performance of the combined interpretations compared to the original interpretation. The alternative interpretations are likely to be scattered across an area close to the original interpretation. Interpretations of ECGs for which the original interpretation differed from alternative ones probably lie close to decision region edges. Alternative interpretations may then lie in a decision region different from that of the original interpretation. If the majority of alternative interpretations lies in a different region, the combined interpretation changes. In cases where the combined interpretation is an improvement, this different region is the correct one. In case of a deterioration, the original interpretation was in the correct region, while the majority of the alternative ones was on the wrong side. This may be caused by, e.g., excessive noise in case of individual beats, or the simulations of electrode displacements not reflecting true displacements for that particular ECG.

Decision trees are popular classifiers because humans are able to read and understand these trees. However, a major drawback is their use of discrete decision thresholds, which makes them sensitive to measurement variability, e.g. intra-individual variability. To decrease the impact of this drawback, careful design of the decision regions is mandatory. The methods presented in this paper may help to accomplish this. First, they can help pinpoint large transitions in the feature space, so that situations where an insignificant change in one measurement causes a large transition in classification can be avoided by correcting the decision tree. Second, they can help to improve a classification by generating alternative classifications, which are subsequently combined. A prerequisite for using these methods is that a model is available to simulate intra-individual variability.

The use of these methods is not restricted to decision trees. They can be applied to any classification algorithm, to provide insight into its stability for intra-individual variability. If a model to simulate intra-individual variability is available, multiple recordings are not necessary.

Conclusion

Intra-individual variability can be employed to improve computerized ECG interpretation. The methods presented do not require multiple recordings and can be used to use the stability index to obtain additional information about the performance of the original interpretation, and to combine alternative interpretations into a more representative interpretation.

The methods presented can be used in these two ways for any classification algorithm and any classification problem where multiple measurements pose a problem, provided a model is available to simulate intra-individual variability.

References

[1] Hurst JW, Treasure CB, Sathavorn CS. Computer errors in electrocardiography. *Clin Cardiol* 1996;19; 580-6.

[2] RuDusky BM. Errors of computer electrocardiography. *Angiology* 1997;48; 1045-50.

[3] Zywietz C, Willems JL, Arnaud P, Van Bemmel JH, Degani R, Macfarlane PW. Stability of computer ECG amplitude measurements in the presence of noise. *Comput Biomed Res* 1990;23; 10-31.

[4] Bailey JJ, Horton M, Itscoitz SB. A method for evaluating computer programs for electrocardiographic interpretation. III. Reproducibility testing and the sources of program errors. *Circulation* 1974;50; 88-93.

[5] De Bruyne MC, Kors JA, Visentin S, Van Herpen G, Hoes AW, Grobbee DE, et al. Reproducibility of computerized ECG measurements and coding in a nonhospitalized elderly population. *J Electrocardiol* 1998;31; 189-95.

[6] Herman MV, Ingram DA, Levy JA, Cook JR, Athans RJ. Variability of electrocardiographic precordial lead placement: a method to improve accuracy and reliability. *Clin Cardiol* 1991;14; 469-76.

[7] Rautaharju PM, Wolf HK, Eifler WJ, Blackburn H. A simple procedure for positioning precordial ECG and VCG electrodes using an electrode locator. *J Electrocardiol* 1976;9; 35-40.

[8] Schijvenaars RJA. *Intra-individual variability of the electrocardiogram*. Thesis. Rotterdam: Erasmus University Rotterdam; 2000.

[9] Schijvenaars RJA, Kors JA, Van Herpen G, Kornreich F, Van Bemmel JH. The effect of electrode position changes on ECG interpretation by computer. *J Electrocardiol* 1997;30; 247-56.

[10] Schijvenaars RJA, Kors JA, Van Herpen G, Van Bemmel JH. Use of the standard 12-lead electrocardiogram to simulate electrode displacements. *J Electrocardiol* 1996;29 Suppl; 5-9.

[11] Van Bemmel JH, Kors JA, Van Herpen G. Methodology of the modular ECG analysis system MEANS. *Methods Inf Med* 1990;29; 346-53.

[12] Willems JL, Abreu-Lima C, Arnaud P, Van Bemmel JH, Brohet C, Degani R, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med* 1991;325; 1767-73.

[13] Michaelis J, Lippold R, Gluck E, Schindler W, Scheidt E. Automated serial ECG analysis within an epidemiological study. *J Electrocardiol* 1987;20 Suppl; 34-6.

Address for correspondence

Bob J.A. Schijvenaars
Department of Medical Informatics
Erasmus University Rotterdam
PO Box 1738
3000 DR Rotterdam
The Netherlands
Email: schijvenaars@mi.fgg.eur.nl.