How a Human Ranks the ECG Diagnostic Parameters: The Pursuit of Experts' Preferences Based on a Hidden Poll

P Augustyniak

AGH University of Science and Technology, Krakow, Poland

Abstract

The paper presents the methodology and results of the research of human expert preferences for particular diagnostic results correlated with the actual patient disease. We applied a custom-build software recording the expert interaction with the computer while he is gathering together all necessary diagnostic statements into a final report. The module works behind the standard ECG interpretation software and recorded 1730 human-verified examinations in 11 cardiologist offices. The statistical processing of these records yield justified parameter hierarchies for 12 most common heart diseases and approximate preferences for other 17. The hierarchy of diagnostic outcome shows his utility in prioritized non-uniform transmission in wireless monitoring networks and in disease-oriented assessment of quality of manual and automated ECG interpretation.

1. Introduction

Despite the common approach considering diagnostic parameters as numeric values of equal relevance, metrological exactness would involve additional attributes as accuracy, time-related validity and adequacy in the assessment of how reliable and useful each data is in description of a particular disease. The last parameter, estimating the contribution of particular diagnostic parameters in final decision is main topic of this paper. The practice shows, that when analysing the numeric outcomes of automated ECG interpretation, human experts apply a certain priority rules in order to verify their hypotheses. Being aware of interpersonal differences, we aimed to reveal and to quantify this hierarchy as far as possible. Our approach puts in question the equal relevance of all diagnostic parameters, usually silently assumed in data processing.

As far, as it can be trusted on the expert skills, the investigation of human preferences yields a new knowledge about the diagnostic data. The data hierarchy itself, and its application in the software is interesting at least for:

- o Attributing data priority for transmission and processing in distributed networks,
- o Assessment of diagnosis quality based on multi-parameter weighted difference,
- o Replicating rules of human reasoning in automatic ECG interpretation software.

Some of these application aspects are currently investigated in AGH-UST Biocybernetic Lab as parts of a project focussed on development of a wearable heart monitor with auto-adaptive interpretation software [1].

2. Methods

2.1. Open and hidden poll research

The usual way of development of new signal processing methods in medicine includes the validation stage performed by human experts. It may have a form of a questionnaire in which the expert expresses his or her opinion in a structured way. This method, although commonly used, has a serious limitation if the experts' own behavior is in question. The expert is hardly able to be self-objective and to describe his action without the reference to the standards. Additionally, the response is affected by memorization and verbalization, two mental processes in a natural way involved in any observation performed by the human. Memorization assigns a considerable part of human attention to capture his own behavior [2], [3], making the behavior not as spontaneous as it is naturally. Moreover the auto-observation usually implies subconscious auto-restriction, resulting in inconsistency and alteration of memorized facts. The willing expression is thus not satisfying a fundamental principle of objective measurements that assumes the less-possible influence of data acquisition process on the observed phenomena.

Hidden poll, although raising ethical questions, is an alternative way of pursuit the experts' knowledge. The only issue is that the human expert being a subject of the experiment is not informed about the recording of his or her behavior. In the metrological aspect, however, it is the principal condition necessary in order to the investigated expert - intentionally or not - does not alter the measurement result. The experts participating in the experiment were informed beforehand that they will be asked for the ECG interpretation and they will be observed at work in some manner. During and after the experiment, the measurement technique remains undisclosed. The experimental setup reproduced as close as possible the natural working environment of the expert.

Hidden poll belongs to behavioral observation techniques and is commonly accepted in social sciences, psychology and medicine. Taking the last domain as an example, the doctor examining his patient usually begins with the interview, but commonly needs supplementary tests providing objective measurements of various diagnostic parameters. The patient without specialized sensors and measurement methodology is not able to estimate several important facts about himself, even assuming he is a highly cooperative proprietor of the information. Following that scheme, we intend to acquire and explore new areas of medical knowledge in order to replicate the human data approach and decision-making path in the software.

2.2. Human expert as experiment subject

The hidden poll was based on commercial ECG interpretation software designed for analysis of typical 12-lead long-term recording from a portable recorder [4]. The interpretation is fully automated except for two stages requiring expert interaction. First is the correctness of heartbeat classification and the second is the selection of diagnostic parameters to be included in the final report. The later stage is where the poll was hidden.

As the cooperating software manufacturer was verifying a new prototype (Cardioteka©, Aspel) in selected cardiology expert offices, we took the opportunity to include the poll subroutine. We applied it in the standard interface of report content selection. In the screen with the proposal of diagnostic report content randomly ordered and pre-selected items were displayed instead of the default screen (fig. 1). Once the interpretation of each ECG is completed, all available report items appear in accidental arrangement, some of them marked as selected. Since the selection is intentionally inappropriate, the doctor had to seek and select (deselect) results he or she wishes to include in (exclude from) the report contents. The order of selections made and chosen items are memorized with the diagnostic outcome.

The randomness of report content selection screen prompts the expert to actively select or deselect report components, and allows us to collect experimental data for studies concerning doctors' preferences. Furthermore, we had to apply a simulated limit of available resources in order to restrain doctors from all-inclusive selections. Each diagnostic parameter was attributed with the value of the expected data stream and the total volume of selected diagnostic data was to be kept below certain level. In such environment, the doctor has to allocate the space in the report first for the most relevant data, and simultaneously exclude the data he or she considers less important (fig. 2).

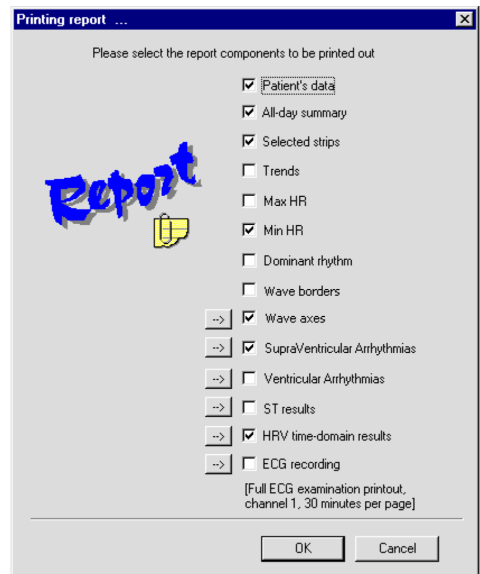

Fig.1. Example selection screen for the choice on the report contents. Subsequent displays differ by items order and initial selection state

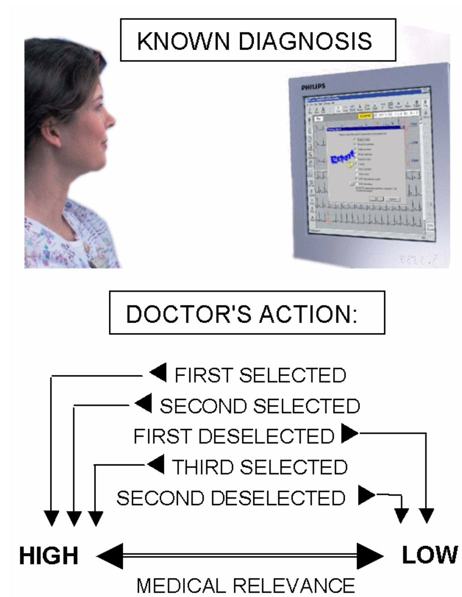

Fig.2. The principle of quantitative doctor's choice measurement leading to disease-specific hierarchy of diagnostic parameters

2.3. Statistical processing of measurements

The statistical processing of the gathered data was aimed at revealing the hierarchy and quantitative relevance of diagnostic parameters on a background of doctors' preferences. Main steps of the calculations were the following:

- Inclusion or exclusion of a parameter to/from a diagnostic report increases or decreases its relevance accordingly to the expert action. First inclusion is the most relevant, first exclusion is the least relevant etc. Items remaining untouched by the expert in the particular examination are not considered for the hierarchy statistics.
- For each considered disease, the disease-specific hierarchy list was build of diagnostic parameters *p*, *p*={1...22} ordered by their frequency *F* of occurrence at a given position L_C relative to all occurrences at other positions *L* weighted by their distances $|L - L_C|$:

$$F = \frac{\sum C : L_C = L}{\sum (C \cdot |L - L_C|) : L_C \neq L} \quad (1)$$

  The maximum *F* for a given *p* indicates the most probable position L_{max} of the parameter in the ranking.
- The diagnostic relevance is represented by the weighting coefficient W_p including the rank L_{max} and the frequency *F*:

$$W_p = \frac{F}{L_{\max}} \quad (2)$$

- Finally, the weighting coefficients W_p were normalized so as they sum to the unity

$$\sum_p W_p = 1 \quad (3)$$

This operation yields a disease-specific vector of weighting coefficients representing the medical relevance of ECG diagnostic parameters.

3. Results

The survey included 1730 ECG analysis cases and allowed to pursuit the cardiologists' preferences in 12 most frequently observed diseases (normal sinus rhythm, sinus tachycardia, sinus bradycardia, probable AV block, ventricular escape beats, atrial fibrillation, AV conduction defect, myocardial infarction, atrial enlargements, ventricular hypertrophies, left bundle branch block, right bundle branch block). The observations count for these pathologies ranged from 16 to 323 cases. For other 17 pathologies, the occurrence frequency was below 16 in the available population and thus no statistically justified conclusions may be drawn from.

The excerpt from the matrix of result attributing to each of the 12 pathologies a hierarchical list of most relevant diagnostic parameters is displayed in Table 1. The table presents normalized values W_p of the weighting coefficients, although due to the lack of completeness the values in rows doesn't sum to the unity. Table 1. presents only first four parameters showing primary relevance throughout the tested examination set. Other 18 parameters has a statistically minor contribution to the report contents, however, in very specific case, they take the priority over those displayed in Table 1.

TABLE I

Disease-specific hierarchy of ECG diagnostic parameters (full matrix contains 22 columns with parameters)

disease	heart rate	dominant trigger src	PQ duration	QRS axis
normal sinus rhythm	0.15	0.25	0.12	0.15
sinus tachycardia	0.55	0.25	0.10	0.13
sinus bradycardia	0.57	0.23	0.10	0.13
probable AV block	0.23	0.17	0.27	0.18
ventricular escape beats	0.27	0.61	0.07	0.05
atrial fibrillation	0.35	0.08	n. a.	0.06
AV conduction defect	0.19	0.13	0.39	0.21
myocardial infarction	0.15	0.12	0.14	0.28
atrial enlargements	0.12	0.02	0.15	0.02
ventricular hypertrophies	0.03	0.07	0.05	0.31
left bundle branch block	0.11	0.18	0.21	0.34
right bundle branch block	0.08	0.17	0.21	0.28

Significant variations of medical relevance may be observed in the results table. In tachycardia, bradycardia and fibrillation the *heart rate* seems to be the most important parameter. In case of ventricular hypertrophies its importance drops dramatically giving priority to other parameters, particularly to QRS axis positions. The *dominant trigger source* shows its primary importance in detection of ventricular escape beats, and also medium importance for distinguishing sinus rhythms from other rhythm types. The *PQ duration* or duration stability, representing the stimulus propagation during an atrioventricular depolarization sequence is a factor of primary relevance in case of AV conduction defects detection, and may be considered as auxiliary determinant in atrioventricular blocks or blocks of His bundle branches. The *electrical axis of QRS complex* represents the ventricular wave front propagation and thus it is much more important in detection of myocardium-related diseases (hypertrophy, infarct) or His bundle problems, than in detection of stimulus triggering or conduction abnormalities.

4. Discussion and conclusions

The aim of presented research was to record and analyze the expert's behavior in order to extract the knowledge about the relative relevance of ECG diagnostic parameters in most frequent diseases. The applied methodology of hidden poll is found to be superior to an open questionnaire investigations because of yielding an unbiased result. Our results confirm the common, but poorly justified belief, that for the human expert some diagnostic results are more important than others. Additionally to the established parameters hierarchy, also the relative relevance factor has been measured and quantitatively expressed. The relevance of particular medical parameters highly depends on the known status of the patient. This confirms that the human expert behavior is driven by hypotheses. The diagnostic parameters hierarchy is applicable as background for automatic software management in systems with patient-specific adaptation of the interpretation processing.

The method involves ethical issues, and probably such methods, although very advantageous, should be used under a supervision of ethical commissions similar to other experiments *in vivo*. The human under test being a proprietor of the knowledge and of the performance, has only a limited influence on the information he or she provides to the analytic system. We carried out our research, in particular at the stage of data acquisition, with maximum care without attempting to judge or compare skills or knowledge of cardiology experts. The method, however, may be considered as a background of human performance assessment. Particular action of the cardiologist under test in response to a purposely prepared set of examination traces may be assessed automatically in an objective way and yield a quantitative estimate of the human expertise.

Similarly to one of the authors' previous works [5], this research demonstrates that automated ECG interpretation systems are far behind the human experts in optimal use of effort and resources. Regular sampling of the signal and uniform consideration of all diagnostic parameters, commonly used today, offer simplicity of procedures at the price of serious unnecessary overload of processors and networks. Investigation of human reasoning and performance is expected to influence the software modification towards a closer simulation of the human interpretation act, not only similar interpretation results.

Acknowledgements

Scientific work supported by the AGH-University of Science and Technology grant No 10.10.120.783. Author expresses his gratitude to the Aspel SA for the consent to software modification and to all cooperating cardiologists for their patience and understanding.

References

[1] Augustyniak P. The use of selected diagnostic parameters as a feedback modifying the ECG interpretation. Proc. Computers in Cardiology; 33, 2006, pp. 825-8

[2] Izawa C. Human learning and memory: Advances in theory and application. Mahwah, NJ: Lawrence Erlbaum Associates. 2005

[3] Mace J. H. (Ed.) Involuntary Memory. Malden, MA: Blackwell Publishing, 2007

[4] IEC 60601-2-51. Medical electrical equipment: Particular requirements for the safety, including essential performance, of ambulatory electrocardiographic systems. 2003,

[5] Augustyniak P. How a Human Perceives the Electrocardiogram. Proc. Computers in Cardiology; 30, 2003, pp. 601-604,

Address for correspondence

Piotr Augustyniak
30, Mickiewicza Ave. 30-059 Krakow Poland
august@agh.edu.pl