Available online at www.sciencedirect.com

ScienceDirect

Journal of Electrocardiology xx (2016) xxx – xxx

JOURNAL OF Electrocardiology

www.jecgonline.com

Review

Digitizing paper electrocardiograms: Status and challenges

George S. Waits, MD, [a,*] Elsayed Z. Soliman, MD, Msc, MS [b]

[a] *Department of Internal Medicine, Wake Forest School of Medicine, Winston Salem, NC*

[b] *Epidemiological Cardiology Research Center (EPICARE), Department of Epidemiology and Prevention, and Department of Internal Medicine, Cardiology Section, Wake Forest School of Medicine, Winston Salem, NC*

Abstract The paper electrocardiogram (ECG) has been widely used for cardiac assessment for well over a century. ECGs can be obtained quickly and cheaply. For this reason, an ever-growing amount of paper ECG records continue to accumulate, some of which are stored into a paper-only format. Converting paper ECGs into digital form has been proposed as the most efficient means to store and analyze an otherwise cumbersome paper archive. In this article, a literature review was conducted for conversion algorithms, criticisms of said algorithms, applications, and standardization efforts. The algorithms were compared in tabulated form. Key functions that have advanced the conversion algorithms as well as remaining challenges are discussed.

Keywords: Digitization; Electrocardiogram; Conversion

Introduction

Ever since Willem Einthoven introduced the electrocardiogram (ECG) as a tool for clinical assessment in 1895, efforts have been made to advance the technology in order to improve the diagnosis and management of cardiovascular disease [1]. The impact of the ECG as a diagnostic tool, as a means to monitor response to treatment, and as a metric for drug effects is well-established. Today, ECGs are routinely recorded, stored, and transmitted in a variety of paper and digital formats. Analyses of these ECGs have permitted the definition of normal and diseased hearts defined by amplitudes, intervals, and morphologies that can be measured. While advances in the digital age enabled large data processing, there remains an expansive, hard-copy ECG data-set that were generated before digital ECG became the norm. It has been hypothesized by many groups that a robust method of converting this unharvested data set of hard-copy ECGs into digital form may significantly expand the understanding of cardiac electrophysiology, the mechanisms of arrhythmias, and their relation to clinical disease [2]. Ready examples of paper ECG stores that could be retrospectively analyzed via digitization include large-scale cardiovascular populations studies such as the REasons for Geographic and Racial Differences in Stroke (REGARDS), The Diabetes Control and Complications Trial (DCCT), the Epidemiology of Diabetes Interventions and Complications (EDIC), and many others stored at the Epidemiological Cardiology Research Center (EPICARE) Reading Center located at Wake Forest School of Medicine, Winston Salem, North Carolina (*personal communication, Soliman EZ, EPICARE Center Director*). This is in addition to paper ECGs generated from preliminary athletic screenings and ECGs in countries lacking the technologic capability of ubiquitous digital storage. The purpose of this paper is to review proposed methods of paper-to-digital ECG conversion, highlight potential applications, and investigate standardization attempts, criticisms, and remaining challenges.

Methods

Search strategy

The phrase “ECG paper digital conversion” yielded 297 results when searching the PubMed Central Database in March 2016. Other query permutations including the terms “EKG,” “electrocardiogram,” “digitized,” “scan,” “hard-copy,” and “converting” were used along with back-referencing in order to gather a broad body of research from which to select relevant papers. From the initial search results, 52 potentially relevant abstracts were selected and reviewed from which 15 were determined to be relevant to the present topic. Four of the relevant papers were found to be unavailable for access. The bibliographies of the remaining 11 papers were reviewed which generated additional sources. Ultimately, 24 papers were included.

* Corresponding author at: Wake Forest School of Medicine, Medical Center Blvd, Winston Salem, NC 27157.

E-mail address: gwaits@wakehealth.edu

http://dx.doi.org/10.1016/j.jelectrocard.2016.09.007
0022-0736/

Table 1
Summary of paper-to-digital ECG algorithms.

First author	Year	Key functions	Output	Validation metrics (statistical tests)	N	Performance results	Software
Zhang [5]	1987	Histogram filtering, line template, isolated-point template; eight neighborhood tracing	1D signal, unspecified	Qualitative only	1	n/a	M75 image system, VAX 11/730 host computer
Widman [7]	1991	Thinning, low-pass filtering, triangle approximation	Bitmap	RR intervals, QRS amplitudes (RMS, Pearson correlation coefficients)	261	"Equivalence" of digital and digitized signals	MS-DOS Turbo C 2.0
Bhullar [3]	1993	Grid filtering, thinning and smoothing, high- and low pass filtering	PCX	RR & QT intervals (correlation coefficient, Bland–Altman, sign tests)	112	Correlation within ~1%	Unspecified graphics editor
Wang [6]	1996	Binarization, thinning, pre- and post-filtering (tracing)	ASCII	Qualitative only	1	n/a	Windows, GUI
Lobodzinski [4]	2003	Optical waveform recognition, waveform segment linking and filtering	XML	Direct comparison (difference plotting)	240	Sample-by-sample differences of 1–3 pixels (4–12 uV)	Windows 2000, MATLAB, Java, C++
Mitra [10]	2004	Binarization, background separation, thinning, data extraction, sorting	DICOM standard, ASCII	Harmonics of V4, V6: normal vs. ischemic vs. infarcted (t test)	60	Accuracy of ~98.4%	Self-developed
Badilini [2]	2005	Grid detection, skew correction, waveform detection, anchor point setting	XML, ASCII, or binary	Direct comparison (RMS), QT intervals (Bland–Altman)	60	Precision within 5–6 ms	Self-developed, C++ based
Karsikas [8]	2007	Signal contour extraction, alignment	ASCII	T-amplitude, T-area, and derived parameters: loop and non-dipolar (Bland–Altman, Median abs. Deviation, Mann–Whitney, RMS error)	30	Mean absolute deviation = +5.85% to −4.86%; relative median error for T wave amplitude = −1.7 ± 5.4%, T wave area = 18.6 ± 21.1%; RMS err: 10–12%	SPSS; "Un-Scan It"
Ravichandran [9]	2013	Grid detection, skew correction, column-wise pixel scanning, optical character recognition	1D signal, unspecified	Direct comparison; PR, QRS, RR, QT, QTc intervals (kappa statistic, signal correlation match)	24	Kappa stat >85%; signal fit 75–80%; correlation 85–90%; intra-observer and inter-observer correlations 80–100%	MATLAB

ASCII - American Std Code for Information Interchange; DICOM - Digital Imaging and Communications in Medicine; GUI - guided user interface; N - number of ECGs sampled; PCX - Picture Exchange; RMS - root mean square; SPSS - Statistical Package for the Social Sciences; XML - extensible markup language;

Inclusion and exclusion criteria

All included papers were written or translated into English with full-text available online or in print through library archives. Papers were included based on relevance to the thesis by satisfying at least one of the four criteria: (1) a unique proposal of a method to convert paper ECGs to digital form, (2) an application of said conversion, (3) a critique of conversion algorithms, or (4) an effort to standardize the conversion process to promote signal fidelity. Excluded studies were either unavailable in an English translation, inaccessible by library licensure limitations, or irrelevant by failing to satisfy the four relevance criteria as listed above.

Quality assessment

No standard qualitative metric was applied as a means of including papers or as a means of rating those papers included. The bibliography was collected by one of the authors (GW) and reviewed by the other (EZS).

Data extraction

Each paper proposing a conversion algorithm was read in full along with the bibliographies. For the purpose of generating a tabular review, the following information was extracted from each of the studies: first author, year, number of ECG samples used, key algorithmic functions, validation metrics, output formatting, supporting software, and claimed outcome. No meta-analysis was performed given the heterogeneity of statistical methods used and outcomes measured. All included papers were queried for proposed limitations of current conversion technology and corresponding solutions offered either explicitly or by implication.

Results

Nine studies were included for presenting unique algorithms aimed at converting paper ECGs into digital form. Three papers were included for raising important criticisms of the conversion proposals. Five papers reflect

Paper ECG
↓
1. Optical Scanning
(±skew correction)
↓
Raw digital ECG with gridline
↓
2. Gridline Removal
↓
Discontinuous, raw ECG without gridline
↓
3. Continuity Algorithms
↓
Continuous raw ECG
↓
4. Additional Processing
↓
Processed digital ECG signal
↓
5. Data extraction
(±optical character recognition)
↓
Digitized ECG

Fig. 1. Generalized schematic of paper-to-digital ECG conversion. Major steps are emboldened. Minor steps are in parentheses. Input and output variables are italicized. Detailed explanations of each step are listed in Table 2.

standardization efforts by groups including the American Heart Association, National Heart Lung and Blood Institute, Common Standards for Quantitative Electrocardiography (aka "CSE Working Party"), the International Council of Harmonization, and the International Society for Computerized Electrocardiology. These groups share a goal of developing a universal method of storing digital ECG information to foster advances in cardiac electrophysiology analysis, diagnostics, and research. Five papers proposed clinically-relevant applications of paper-to-digital conversion technology, and an additional two papers were included to provide historical context.

Table 1 summarizes the studies proposing a method of converting paper ECGs into digital form. Publication years ranged from 1987 to 2013. The number of paper ECG samples used ranged from 1 to 261 samples. Output data format varied with the American Standard Code for Information Interchange (ASCII) occurring most frequently. Supportive software ranged widely with only MATLAB and Windows being explicitly referenced more than once.

Fig. 1 displays the order of key functions involved in the paper-to-digital conversion for each algorithm. Table 2 includes a comprehensive list of each function along with a brief description of the step. All algorithms began with a scanning process, some applying a color-filter prior to the scan with intent to remove the gridline [3], others using line size or gray-scale value to make the distinction. The algorithms would then proceed to generate a digital ECG signal through some permutation of functions as listed in Fig. 1. Some notable steps include "skew correction" – to account for rotated scans, "thinning" – the process of reducing the inherent, non-uniform thickness of an ECG tracing as rendered by the ink of a mechanical ECG spray writer into a function with a single y-value (voltage) for every x-value (time), various tracing algorithms ("anchor-point setting," "column-wise pixel scan," "eight-neighbor tracing," "signal contour extraction," "smoothing," "triangle-approximation," and "waveform segment linking and filtering"), and additional post-processing such as high- and low-pass filtering. Some algorithms also included a means of extracting demographic data [4].

Validation was attempted using a variety of approaches. Most studies used at least one quantitative validation metric while two relied only on qualitative visual comparison [5,6]. For studies using a quantitative approach, often a printed ECG was digitized via the proposed algorithm and compared against a gold-standard analog-to-digital signal generated from the same recording. Most validation designs included comparison of intervals (e.g., PR, QRS, RR, QTc), amplitudes, areas, or a whole-scale, signal overlay comparison in which overall signal differences were calculated [2–4,7–9]. A few studies compared derived parameters such as the frequency-based "harmonics" calculated using Fourier transform, or trigonometric properties of angles between wave axes [8,10]. Statistical tests used to compare these differences included: root mean squares, Pearson correlation coefficients, Bland–Altman tests (usually for QT intervals), sign tests, t-tests, median absolute deviation, Mann–Whitney tests, and kappa statistics (see Table 1).

Outcomes of each algorithm were reported according to the validation metric used, if any. The highest accuracy claims appeared to come from Bhullar et al. who reported correlation to "within ~1%" and Ravichandran et al. who calculated a kappa statistic of 1.00 for inter-observer measurement of waveform intervals. Conversely, Ravichandran et al. also included the least accurate results in other metrics, reporting a 75–80% "signal fit" match, and Karsikas et al. reported a RMS error of 10–12%. One study's accuracy claim was made in clinically significant terms, reporting that the parameters that originally had discriminating power between healthy and diseased subjects retained their difference after the digitizing process [8]. Whether method accuracy was independently associated with any of the other study parameters – publication year, software used, number of ECG samples processed, key functions included, or the sequence in which these functions were performed – could not be determined due to the heterogeneity of accuracy assessments between algorithms.

Table 3 lists the challenges and corresponding proposed solutions, if any, to the process of converting paper ECGs into digital form. The list is a product of extraction from the discussion sections of the algorithm papers as well as the other papers written for criticism, standardization, and application proposal.

Solutions were aimed at promoting signal accuracy, time-efficiency, clinical-utility, and expansion of quantitative ECG research. Not all limitations listed in Table 3 included an explicit corresponding solution. Some limitations were without corresponding solutions proposed for reasons that were self-evident. For example, that a digitized ECG would always be limited by the errors inherent to the

Table 2
Conversion steps explained.

Conversion step	Brief description	Reference
Optical scanning	The routine process of digitally registering a paper image into digital form.	all
Skew correction	The process of selecting two points along the same gridline, calculating the slope angle and rotating the image to eliminate the skewed angle.	Ravichandran [9], Badilini [2], Lobodzinski [4]
Grid removal (a.k.a., "histogram filtering," "grid filtering," "grid detection," "pre-filtering," "binarization", "edge-detection")	The process of separating the ECG signal from the paper grid (using color, thickness, gray-scale value differences or other means of distinction). Sometimes using 1xN masks to scan across the pre-filtered image for noise reduction, other times using histogram-based assays.	Zhang [5], Bhullar [3], Badilini [2], Ravichandran [9], Lobodzinski [4], Wang [6], Mitra [10]
Continuity algorithms (e.g., "line template," "isolated-point template," "eight-neighbor tracing," "smoothing," "triangle approximation," "anchor-point setting," "post-filtering," "waveform segment linking," "thinning," "waveform detection," "signal contour extraction," "column-wise pixel scan")	Various algorithms used to detect the ECG signal, distinguish it from surrounding noise and gridline, and generate a continuous ECG signal through interpolation. The result is a continuous ECG signal with some superimposed noise that may require additional processing. For example, Eight-neighbor tracing horizontally scans a 9 × 9 matrix beginning with a known centered point on the ECG signal.	Zhang [5], Bhullar [3], Wang [6], Lobodzinski [4], Widman [7], Mitra [10], Badilini [2], Karsikas [8], Ravichandran [9]
Additional processing (e.g., "high- and low-pass filtering," "alignment")	Various processing steps performed on a continuous ECG signal to minimize the noise and improve the accuracy of the ECG parameters.	Widman [7], Bhullar [3], Karsikas [8]
Data extraction (a.k.a., sorting)	The process of extracting the digitized filtered ECG into a storage format.	all
Optical character recognition	The process of scanning and cross-referencing printed text with a pre-defined character template database and facilitating storage of demographic data along with the ECG signal itself.	Ravichandran [9], Lobodzinski [4]

original paper ECG. Other solutions may have been omitted due to scope limitations (e.g., the authors would not report every combination of ECG writer and paper type to propose an ideal combination that minimizes the error of "graininess" due to decreased ink retention). References to limitations raised and corresponding solutions, if present, were also included.

Discussion

This study sought to summarize the methods proposed for generating digitized ECGs from paper tracings. A number of studies underscore the importance of these conversion efforts [11–14].

Efforts to digitize paper ECGs are largely driven by the greater aim of maximizing the collective ECG record to promote research. Therefore, as part of the digitizing process, many have recognized the need for a uniform format. Historically, there have been numerous efforts toward this goal. Kothadia et al. recognized the conflict between the patient-centered research interest of format standardization and the economic interest of ECG manufacturers to maintain proprietary formats not directly compatible with other ECG manufacturers. Kothadia attributes the failure of standardization efforts by Willems et al. in the 1990s and Bailey et al. in the 1970s to this competing economic interest [15–17]. Kothadia further noted the labor-intensive and often redundant efforts made earlier by Miyahara in 1984 involving transcription of a proprietarily-digitized signal backwards into analog via magnetic tape before re-constructing the signal into a more usable digital form [18]. As perhaps an improvement on these previous efforts, the development of rapid algorithms for paper-to-digital conversion could introduce a means of achieving a universal ECG format. In other words, if the numerous stores of digital ECG recordings existing in various formats cannot be reconciled into a single record due to proprietary firewalls, a feasible alternative may be to combine them using their visual (i.e., printed) presentations.

One case-in-point for the digitization effort is noted in Badilini's ECGScan algorithm [2]. The software developed in the Badilini method required a user to drag an "active rectangle" around the scanned ECG signal of interest. Within the rectangular selection, the software would detect, process, and generate a digitized ECG signal using mathematical assumptions. In Badilini's algorithm, he utilized a "cost function," defined as the weighted mathematical sum of

Table 3
Enumeration of limitations and solutions by reference.

Limitation	Solution	Reference
Grid issues:		
Identifying the grid	Various grid modes: 1. Exact mode – input parameters are given (paper speed, scanning resolution, distance between gridlines) to determine the number of pixels between gridlines 2. Range mode – similar to "exact mode," but adds a confidence level to accommodate poorer quality ECGs (faxed, photocopies) 3. Manual mode – uses no input parameters, requires user to draw an "active rectangle" around a clear grid sample of 4–5 squares	Badilini [2]
Inherent error in printing ("error lost in printing is forever lost")	Optimizing machinery (electrocardiograph), materials (grid paper), and operator performance[1]	Stockbridge [11], Baeza [12]
Gridline removal	1) "Background separation" used to exclude the presumed dotted gridlines by setting a distance threshold between two interrogated points 2) Removal based on line-width difference between signal and grid 3) Color-filter scanning method	1) Mitra [10] 2) Wang [6] 3) Bhullar [3]
Facsimiles distort intervals	"Other modes" of transmission	Farooqi [24]
Time efficiency:		
User-interaction required; lack of full automaticity	Function permitting multi-lead analysis simultaneously (e.g., the user draws a single "active rectangle" around multiple leads, processed simultaneously	Badilini [2]
Time-consuming	Automaticity, development of a standard conversion method, minimize processing workload	Baeza [12], Ravichandran [9]
Alignment issues:		
Loss of relative timing between leads	Multi-lead analysis functionality as above	Badilini [2]
Need for time alignment when using older ECG devices (which separate precordial and limb leads with unspecified time shift)	Multi-lead analysis, alignment algorithm based on a user-selected QRS start point. Found to be more accurate than the more automatic R-wave peak wave detection approach	Badilini [2], Karsikas [8]
Signal quality/noise issues:		
Inaccuracy in high frequency regions	Triangle approximation method: user selects points which are incorporated into the automated conversion. Also found adjusting paper speed could prevent upstroke/downstroke overlap	Widman [7]
Noise	Beat averaging, minimum wave thresholding, re-labeling QRS complex to optimize alignment	Zywietz [21]
Salt-and-pepper noise from gridline-removal	1) 3 × 3 pixel median filtering, and interpolation 2) Eight-neighbor tracing filter	1) Ravichandran [9] 2) Zhang [5]
Sources of noise to ECG: power lines (50 Hz), EMG, surgical noise, respiration, baseline shifting	Test environment optimization[1]	Mitra [10]
High frequency noise (e.g., due to non-uniform ink retention)	"Smoothing" function, low-pass filtering	Bhullar [3]
Interval distortion from heterogeneous signal thickness	Thinning algorithms Example: thicknesses defined by number of pixels, middle pixel selected, adjacent pixels connect to form a continuous signal with a single-pixel width	Various; example from Widman [7]
Validation issues:		
Failure to demonstrate robust conversion of large, wide-ranging ECG tracings representing the general population	Testing on at least one hundred samples	Lobodzinski [4], Widman [7], and Bhullar [3]

[1] Solution derived implicitly.

"line," "smoothness," and "length" functions. This method was praised in an editorial by Richard Baeza in 2005 who found the method particularly useful for its versatility in digitizing both single- and multi-lead ECG signals [12]. On the other hand, Baeza critiques that the Badilini method makes no claim regarding the time required to scan, process, and generate the digital signal. It was also cautioned that any digitizing algorithm, regardless of its robustness, would always be limited by the quality of the printed ECG upon which it acts. Additional related criticisms included the omission of an explicit description of the scanning procedure. Baeza also questioned whether a technique tested on only 60 ECG samples could demonstrate the versatility necessary to process the wide varieties of electrocardiographically-distinct tracings of normal and diseased subjects found in the general population [12]. However, in a separate editorial, Stockbridge praised Badilini's XML output as meeting standards set by the FDA for purposes of demonstrating drug effects on QT prolongation, while also offering criticism of the ECGScan method [11]. Specifically, he noted the labor-intensity of the method which required a user to manually draw "active rectangles" over the signal of interest rather than a more automated system. He also identified the technical difficulty of maintaining signal fidelity, especially when processing high frequencies [11].

Hingorani et al. conducted a thorough analysis of Badilini's ECGScan program with the goal of identifying the respective error contributed by the printing, scanning and digitization processes [19]. This study methodically generated 2 ECGs from each of 50 volunteers using 2 ECG machines. Each machine would print the recorded ECG twice, each print was scanned twice, and each scanned ECG was digitized twice using Badilini's software. Printer speed was found to vary by ~5% while the scanning process produced negligible error in keeping with scanning standards requiring tolerance to 1%. The digitalized ECG itself was found to contribute the greatest amount of error as compared with the gold standard digital ECG that was generated directly from the ECG machine. With a clear negative judgment, Hingorani et al. concluded that the digitization software failed to meet the repeatability standards as required by the International Conference on Harmonization (ICH) for the purpose of assessing drug effects on QT/QTc intervals [20]. Without outlining a path for improvement, Hingorani et al. made that case that progress was still needed. It is worth noting that Badilini's algorithm has undergone at least two version upgrades, the most recent released in 2014. The latest version is reported to use new image processing tools to improve the speed and accuracy of the conversion algorithm but it remains to be proven whether the latest version would meet the ICH minimum standards.

Later conversion methods have been proposed including those by Karsikas and Ravichandran [8,9]. Karsikas employed a highly-quantitative approach, whereby he compared parameters involving T-wave and QRS-areas, morphologies, and "non-dipolar" angles-between-axes to assess the fidelity of the digitized paper ECG as compared with a directly-digital gold standard. He also investigated the effect of non-simultaneous lead alignment. Karsikas concluded digitized T-wave parameters maintained robustness and ability to discriminate between healthy and diseased subjects while QRS parameters were found lacking as compared with the directly-to-digital gold standard, perhaps owing to the relative increased signal-to-noise ratio of the T-wave [8]. While the Karsikas method seems to demonstrate an objective, mathematically-based approach; it is worth noting that printed ECGs were made on blank – not standard – grid paper, raising questions about the conversion algorithm's applicability for use with actual paper ECGs.

In Ravichandran's 2013 MATLAB-based approach, he began with scanning, scaling, and skew-correcting using gridline contours to generate an extracted ECG signal. Next, a binary image of the signal was generated using a thresholding technique followed by median filtering and interpolation to eliminate the "salt-and-pepper" noise generated by thresholding. An optional smoothing step was also employed to reduce high-frequency noise. Validation was argued using a three-step approach. First, the digitized paper ECG was plotted over a raw (i.e., directly-digitized) signal and a kappa statistic was calculated to show fitness. Second, ECGs printed at a sample rate of 150 Hz were converted into 1D digitized signals and compared with the 500 Hz raw signal down-sampled to 150 Hz. Finally, clinically significant intervals (PR, QRS, RR, QT, QTc) from various leads where calculated and compared. In all but the direct fit approach, kappa statistics were found to be greater than 0.8 representing excellent agreement. Ravichandran also employed a novel "Optical Character Recognition" feature whereby demographic meta-data could be identified, extracted, and included from the paper record into the digital conversion which adds a clinically-useful improvement to the Badilini and Karsikas methods [9].

Key functions

To effect the needed improvement in the digitization process, some key functions were designed, organized in Table 1 and explained in Table 2. One initial challenge involved the process of gridline removal. Some methods sought to take advantage of the color difference between signal (usually black) and gridline (often red) by applying a color filter into a film during the scanning step [3]. Others used histogram-based filters to take advantage of differences in line thickness, assuming the gridline to have a significantly shorter line-thickness than the signal [5,10]. No method resulted in a flawless gridline removal. Ravichandran et al. used the term "salt-and-pepper" noise to describe the effect of gridline removal with their histogram-based method [9]. This resulting noise would require additional processing approached differently by the various designers. Bhullar used high- and low-pass filtering on thinned images and showed good reproducibility by comparing signals redundantly digitized from the same paper ECG [3]. Widman showed a novel approach to maintaining signal fidelity in the error-prone high frequency regions of the ECG signal (e.g., the QRS complex) by developing a "triangle approximation method," which showed some success but was not entirely automated [7]. Ravichandran's group proved that processing time could be saved by showing insignificant differences between signals generated from 150 Hz sample rates and the larger 500 Hz sample rates. Similarly, they showed no benefit using a higher-resolution scanner, noting that a 300dpi scanner compared equivalently with the 600dpi resolution. Identifying minimum specifications for accurate conversion laid the groundwork for faster algorithms. The need to reduce the processing time is clear when compared to the estimates reported by Bhullar et al. in 1993 of a 20–60 min processing time to digitize a single ECG followed by 2 min for each interval analysis [3]. Through continued efforts and various algorithms as illustrated in Fig. 1, optimism for developing an accurate digitization process began to grow. In fact, Widman et al. claimed that their method satisfied accuracy standards as outlined by the American Heart Association [7,14]. Techniques to correct skew, automatically import demographic data, smoothing, filtering, and aligning were all employed to promote the quality of the digitized ECG signal. Once achieving quality, the next collective task would be to attain uniformity.

Standardization

Zywietz et al. noted efforts at ECG standardization dating back to 1938 [21]. Kligfield reported efforts by the International Society of Computerized Electrocardiology (ISCE) in 1992 to express interest in universalizing the ECG recording format, coining an "ECG Genome Project," though no specific modality was proposed at that time [22]. The ISCE effort was ultimately superseded by introduction of the FDA-mandated XML format [23]. Mitra's method suggested that achieving an ECG standard form could promote data transfer to facilitate patient care, drawing attention to the example of the well-known DICOM standard used by radiology in image data communication [10]. Some groups have been more hesitant to push for uniformity, with only a minority of the National Heart, Lung, and Blood Institute's workshop committee advocating for a collective ECG registry [13]. Barriers to standardization have been attributed to the proprietary interests of the independent electrocardiograph manufacturers [15].

Application

The potential utility of a robust ECG conversion method is broad. In addition to making available the large stores of old paper ECGs for epidemiologic study (e.g., old epidemiologic study records, athletic and insurance screenings, and records from developing countries), other applications are possible. Farooqi experimented with transmission of ECG data via fax machine but found that temporal intervals were significantly distorted [24]. This finding drew attention to the need to develop transmission methods that better preserved the signal. Lin showed that digitization enabled the extraction of buried P-waves using a novel QRST subtraction technique. This method resulted in an average 85% comparative accuracy between the extracted and the reference p-wave of interest [25]. The admitted drawback to the method was the requirement for manual alignment, which could make the method too laborious for practical use. These are just two examples showing the rapid transmissibility and analytics made possible by digital conversion of the ECG.

Conclusions

For over a century, the electrocardiogram has served as a key tool in the disease and management of certain cardiovascular illnesses. Digitization of the large data sets of hard-copy ECGs has promise to expand the study of arrhythmias and cardiovascular disease trends over decades. Numerous algorithms have been proposed for achieving this conversion task and certain limitations have been identified and overcome while others remain elusive. Many see the broader goal of ECG digitization as a means for creating a more universal ECG record to promote rapid transmission, research advancement, and better analytics in order to improve patient care.

Disclosures

None.

References

[1] Barold SS. Willem Einthoven and the birth of clinical electrocardiography a hundred years ago. Card Electrophysiol Rev 2003;7:99–104.

[2] Badilini F, Erdem T, Zareba W, Moss AJ. ECGScan: a method for conversion of paper electrocardiographic printouts to digital electrocardiographic files. J Electrocardiol 2005;38:310–8, http://dx.doi.org/10.1016/j.jelectrocard.2005.04.003.

[3] Bhullar HK, Fothergill JC, Goddard WP, de Bono DP. Automated measurement of QT interval dispersion from hard-copy ECGs. J Electrocardiol 1993;26:321–31.

[4] Lobodzinski SM, Teppner U, Laks M. State of the art techniques for preservation and reuse of hard copy electrocardiograms. J Electrocardiol 2003(36 Suppl):151–5.

[5] Zhang Z, Zhang H, Zhuang T. One-dimensional signal extraction of paper-written ECG imagen and its archiving. SPIE Vis Commun image process II; 1987. p. 419–22.

[6] Wang JT, Mital DP. A microcomputer-based prototype for ECG paper record conversion. J Netw Comput Appl 1996;19:295–307, http://dx.doi.org/10.1006/jnca.1996.0020.

[7] Widman LE, Hines LS. Digitization of electrocardiograms by desktop optical scanner. J Electrocardiol 1991;24:325–38.

[8] Karsikas M, Huikuri H, Perkiömäki JS, Lehtola L, Seppänen T. Influence of paper electrocardiogram digitizing on T wave and QRS complex morphology parameters. Ann Noninvasive Electrocardiol 2007;12:282–90, http://dx.doi.org/10.1111/j.1542-474X.2007.00176.x.

[9] Ravichandran L, Harless C, Shah AJ, Wick CA, Mcclellan JH, Tridandapani S. Novel tool for complete digitization of paper electrocardiography data. Transl Eng Health Med 2013;1, http://dx.doi.org/10.1109/JTEHM.2013.2262024 [1800107–1800107].

[10] Mitra S, Mitra M, Chaudhuri BB. Generation of digital time database from paper ECG records and Fourier transform-based analysis for disease identification. Comput Biol Med 2004;34:551–60, http://dx.doi.org/10.1016/j.compbiomed.2003.08.001.

[11] Stockbridge N. Points to consider in electrocardiogram waveform extraction. J Electrocardiol 2005;38:319–20, http://dx.doi.org/10.1016/j.jelectrocard.2005.06.090.

[12] Baeza R, Wildman LE, Hines LS, Mitra S, Mitra M, Chaudhuri BB. Electrocardiogram digitization: a practical perspective on the usefulness of a new tool to convert paper electrocardiograms into digital waveform. J Electrocardiol 2005;38:321–3, http://dx.doi.org/10.1016/j.jelectrocard.2005.06.100.

[13] Norman JE, Bailey JJ, Berson AS, Haisty WK, Levy D, Macfarlane PM, et al. NHLBI workshop on the utilization of ECG databases: preservation and use of existing ECG databases and development of future resources. J Electrocardiol 1998;31:83–9.

[14] working partyEur Heart J 1985;6:815–25.

[15] Kothadia R, Kulecz WB, Kofman IS, Black AJ, Grier JW, Schlegel TT, et al. New system for digital to analog transformation and reconstruction of 12-lead ECGs. PLoS One 2013;8e61076, http://dx.doi.org/10.1371/journal.pone.0061076.

[16] Bailey JJ, Horton M, Itscoitz SB. A method for evaluating computer programs for electrocardiographic interpretation. 3. Reproducibility testing and the sources of program errors. Circulation 1974;50:88–93.

[17] 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. Med 1991;325:1767–73, http://dx.doi.org/10.1056/NEJM199112193252503.

[18] Miyahara H, Domae A, Sato T. The reproducibility of interpretation of 10 computer ECG systems by means of a microprocessor-based ECG signal generator. Comput Biomed Res 1984;17:311–25.

[19] Hingorani P, Karnad DR, Panicker GK, Deshmukh S, Kothari S, Narula D, et al. Differences between QT and RR intervals in digital and digitized paper electrocardiograms: contribution of the printer, scanner, and digitization process. J Electrocardiol 2008;41:370–5, http://dx.doi.org/10.1016/j.jelectrocard.2008.02.026.

[20] Food and Drug Administration, HHS. International conference on harmonisation; guidance on E14 clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs; availability. Notice. Fed Regist 2005;70:61134–5.

[21] Zywietz C, Willems JL, Arnaud P, van Bemmel JH, Degani R, Macfarlane PW. Stability of computer ECG amplitude measurements in the presence of noise. The CSE working party. Comput Biomed Res 1990;23:10–31.

[22] Kligfield P. Overview of the ISCE ECG "genome project. J Electrocardiol 2003(36 Suppl):163–5.

[23] Stockbridge N. FDA XML data format revision B requirements specification FDA XML data format requirements specification; 2002.

[24] Farooqi KM, Ceresnak SR, Freeman K, Pass RH. Electrocardiograms transmitted via facsimile may not allow accurate interval interpretation. PACE –. Pacing Clin Electrophysiol 2011;34:1283–7, http://dx.doi.org/10.1111/j.1540-8159.2011.03158.x.

[25] Lin B, Wang PJ, Mahapatra S, Homoud M, Link M, Estes NAM, et al. Extraction of buried P waves from printed electrocardiograms. Ann Noninvasive Electrocardiol 2005;10:142–5, http://dx.doi.org/10.1111/j.1542-474X.2005.05605.x.