CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

UCC Library and UCC researchers have made this item openly available. Please let us know how this has helped you. Thanks!

Title	A deep neural network for 12-lead electrocardiogram interpretation outperforms a conventional algorithm, and its physician overread, in the diagnosis of atrial fibrillation
Author(s)	Smith, Stephen W.; Rapin, Jeremy; Li, Jia; Fleureau, Yann; Fennell, William; Walsh, Brooks M.; Rosier, Arnaud; Fiorina, Laurent; Gardella, Christophe
Publication date	2019-09-08
Original citation	Smith, S. W., Rapin, J., Li, J., Fleureau, Y., Fennell, W., Walsh, B. M., Rosier, A., Fiorina, L. and Gardella, C. (2019) 'A deep neural network for 12-lead electrocardiogram interpretation outperforms a conventional algorithm, and its physician overread, in the diagnosis of atrial fibrillation', IJC Heart & Vasculature, 25,100423 (6pp.). DOI: 10.1016/j.ijcha.2019.100423
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://www.sciencedirect.com/science/article/pii/S2352906719301241?via%3Dihub http://dx.doi.org/10.1016/j.ijcha.2019.100423 Access to the full text of the published version may require a subscription.
Rights	**©2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).Contents lists available atScienceDirectIJC Heart & Vasculaturejournal homepage:http://www.journals.elsevier.com/ijc-heart-and-vasculature** http://creativecommons.org/licenses/by-nc-nd/4.0/
Item downloaded from	http://hdl.handle.net/10468/8771

Downloaded on 2021-11-27T10:38:22Z

UCC University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

IJC Heart & Vasculature 25 (2019) 100423

Contents lists available at ScienceDirect

IJC Heart & Vasculature

journal homepage: http://www.journals.elsevier.com/ijc-heart-and-vasculature

A deep neural network for 12-lead electrocardiogram interpretation outperforms a conventional algorithm, and its physician overread, in the diagnosis of atrial fibrillation

Stephen W. Smith [a,*,1], Jeremy Rapin [b,2], Jia Li [b,3], Yann Fleureau [b,4], William Fennell [c,5], Brooks M. Walsh [d,6], Arnaud Rosier [e,7], Laurent Fiorina [f,8], Christophe Gardella [b,9]

[a] Hennepin County Medical Center, Department of Emergency Medicine, University of Minnesota, United States of America
[b] Cardiologs® Technologies, Paris, France
[c] University College, Cork, Ireland
[d] Dept of Emergency Medicine, Bridgeport Hospital, Bridgeport, CT, United States of America
[e] Service de rythmologie, Hôpital privé Jacques Cartier, Groupe GDS, Massy, France
[f] Hopital privé Jacques Cartier, Massy, France

ARTICLE INFO

Article history:
Received 3 June 2019
Received in revised form 20 August 2019
Accepted 2 September 2019
Available online xxxx

Keywords:
Deep neural network
Artificial intelligence
Atrial fibrillation
Atrial dysrhythmia
Electrocardiogram

ABSTRACT

Background: Automated electrocardiogram (ECG) interpretations may be erroneous, and lead to erroneous overreads, including for atrial fibrillation (AF). We compared the accuracy of the first version of a new deep neural network 12-Lead ECG algorithm (Cardiologs®) to the conventional Veritas algorithm in interpretation of AF.
Methods: 24,123 consecutive 12-lead ECGs recorded over 6 months were interpreted by 1) the Veritas® algorithm, 2) physicians who overread Veritas® (Veritas® + physician), and 3) Cardiologs® algorithm. We randomly selected 500 out of 858 ECGs with a diagnosis of AF according to either algorithm, then compared the algorithms' interpretations, and Veritas® + physician, with expert interpretation. To assess sensitivity for AF, we analyzed a separate database of 1473 randomly selected ECGs interpreted by both algorithms and by blinded experts.
Results: Among the 500 ECGs selected, 399 had a final classification of AF; 101 (20.2%) had ≥1 false positive automated interpretation. Accuracy of Cardiologs® (91.2%; CI: 82.4–94.4) was higher than Veritas® (80.2%; CI: 76.5–83.5) ($p < 0.0001$), and equal to Veritas® + physician (90.0%, CI:87.1–92.3) ($p = 0.12$). When Veritas® was incorrect, accuracy of Veritas® + physician was only 62% (CI 52–71); among those ECGs, Cardiologs® accuracy was 90% (CI: 82–94; $p < 0.0001$). The second database had 39 AF cases; sensitivity was 92% vs. 87% ($p = 0.46$) and specificity was 99.5% vs. 98.7% ($p = 0.03$) for Cardiologs® and Veritas® respectively.
Conclusion: Cardiologs® 12-lead ECG algorithm improves the interpretation of atrial fibrillation.
© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Computer electrocardiogram (ECG) interpretation algorithms aim to improve physician ECG Interpretation, reduce medical error, and expedite patient care. Automated interpretations may be overread by the health care provider (HCP) and corrected if necessary. Among rhythm diagnoses, atrial fibrillation (AF) is particularly important for appropriate management. However, incorrect automated and physician overread interpretations are common and have been shown to adversely affect patient management [1]. A study of 2298 ECGs from 1085 patients which had a computerized interpretation of AF found that in 442 (19%) of these ECGs, from 382 patients (35%), the interpretation was incorrect, and that, in 92 of these 382 patients, the physician had failed to correct it. These errors resulted in unnecessary antiarrhythmic and anticoagulant therapy in 39 patients and unnecessary

Abbreviations: AF, atrial fibrillation; AFL, atrial flutter; AT, atrial tachycardia; AD, atrial dysrhythmia; ECG, electrocardiogram; ED, emergency department; DNN, deep neural network; HCP, health care provider.

* Corresponding author at: Department of Emergency Medicine, Hennepin County Medical Center, 715 Park Ave., ER R-2, Minneapolis, MN 55415, United States of America.
E-mail address: smith253@umn.edu (S.W. Smith).

[1] SWS conceived the study, designed the study, provided the ECGs, supervised the conduct of the trial and data collection, assisted in data analysis, and wrote the manuscript.
[2] JR was co-writer of the first version of the deep neural network algorithm, organized the study, managed the data, analyzed the data, and assisted with manuscript preparation.
[3] JL was co-writer of the first version of the deep neural network algorithm.
[4] YF was co-writer of the first version of the deep neural network algorithm.
[5] Expert Interpreter. WF assisted with manuscript preparation.
[6] BW was expert interpreter and assisted with manuscript preparation.
[7] AR was expert interpreter.
[8] LF provided the final decision on ECG rhythm diagnosis.
[9] CG managed the data, analyzed the data, and assisted with manuscript preparation.

https://doi.org/10.1016/j.ijcha.2019.100423
2352-9067/© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

diagnostic testing in 90 patients, and an incorrect final diagnosis of paroxysmal AF in 43 patients.

Deep neural networks (DNN) have proven successful for multiple types of diagnosis, including skin cancer [2,3], diabetic retinopathy [4], computed tomography of the head [5], pneumonia on chest radiograph [6], and detection of lung cancer on chest computed tomography [7].

Recently, Smith et al. [8] reported on a convolutional deep neural network (Cardiologs®), able to detect 76 pathologies simultaneously from 12-lead ECGs (Fig. 1). This algorithm has already proven more specific and accurate, with much higher positive predictive value, than the Veritas® conventional algorithm, for overall emergency department (ED) ECG interpretation; it was also more sensitive for less difficult ECGs. Here we sought to evaluate the accuracy of Cardiologs® DNN Algorithm in the interpretation of atrial fibrillation, along with other atrial dysrhythmias (AD), including atrial tachycardia (AT) and atrial flutter (AFL), in a cohort of hospitalized and/or ED patients. We compared Cardiologs® to both the automated Veritas® diagnostic, and to the physician overread.

2. Methods

The study was approved by the Institutional Review Board of Minneapolis Medical Research Foundation at Hennepin County Medical Center (HCMC) in Minneapolis, MN, USA, IRB number 15-4100. All ECGs were de-identified at the source.

2.1. Selection of ECGs

24,123 consecutive 12 lead ECGs recorded in any hospital department over 6 months at one institution in 2016 were selected. As part of clinical care, and at the time of recording, all ECGs, which were recorded on a Mortara® device, were given an automated interpretation by the embedded Veritas® algorithm (Veritas®). This interpretation was permanently recorded. All ECGs had a physician overread by either an emergency physician or cardiologist; this interpretation was also permanently recorded (Veritas® + physician). The Cardiologs® algorithm was applied to all ECGs; Cardiologs® interpretation was recorded (Cardiologs®).

We randomly selected 500 out of 858 ECGs that were diagnosed with an atrial dysrhythmia (AD) for either Veritas® or Cardiologs. Diagnostic labels were considered AD regardless of ventricular response or atrioventricular conduction. If the algorithm label was "atrial fibrillation or flutter," the rhythm was classified for the study as AF, since this would prompt the physician to scrutinize the ECG. Additionally, for the secondary outcome, all ECGs with a label of AF, AFL, or AT by either or both of the 2 algorithms were selected for study. We limited our analysis to cases with positive automated diagnosis, instead of analysing all 24,123 cases, because if the Veritas® algorithm had not made any diagnosis of AF, AT or AFL, the physician overread was often in free text and therefore difficult to search by automated methods, and too laborious to search manually. For this reason, true overall sensitivity of the algorithms in this dataset could not be studied; sensitivity was calculated only among the cases identified as positive by one or both algorithms.

Each of the 500 ECGs was reviewed by 2 of 3 expert physicians blinded to each other and to all other interpretations. These ECGs were given a specific atrial rhythm diagnosis. For discrepancies, another blinded expert [an electrophysiologist, (LF)] was the tiebreaker, which was the reference standard. Final diagnoses were AF vs. not-AF, and AD vs. not-AD. The primary and secondary outcomes, respectively, were the comparisons of the interpretations by Veritas® and Cardiologs® algorithms, and final clinical overread (Veritas® + physician), with the expert reference standard.

In order to estimate overall true sensitivity, we also queried a database of 1500 ECGs that were randomly selected from all the ED ECGs recorded during the same time period, and used for a previous study [8]. Because 27 of these ECGs were deemed too noisy for expert interpretation (even though the algorithm had provided an interpretation), 1473 had expert interpretation for a reference standard. However, for these 1473 ECGs, the final clinical physician overread (Veritas® + physician) was not available.

Statistics were by percentages with confidence intervals (CI), and with two-tailed Chi Square, and Fisher exact test.

3. Results

Over 6 months in 2016, there were 24,123 ECGs recorded at HCMC. One or both of the algorithms diagnosed AF in 858 ECGs; 500 ECGs were randomly selected. 399 (79.8%) had a final expert classification of AF, and 414 (82.8%) of AD (see Table 1 for the adjudicated diagnoses). There were 101 false positive AF diagnoses by one or both algorithms for AF, and 86 for AD. The initial 2 experts agreed on AF for 471 out of 500 (94.2%), and on AD for 474 out of 500; thus, 29 and 26 needed tiebreaking. For AF, the discordance with the reference was comparable among individual interpreters (96.7%, 96.8%, and 97.6%).

3.1. Cardiologs® compared to Veritas® or to Veritas® + physician

See Table 2 for the primary outcome of AF vs. no-AF. There were 101 (20.2%) ECGs with at least 1 false positive automated interpretation of AF; 10 were misdiagnosed by both algorithms and 91 were misdiagnosed by only one algorithm: 23 by Cardiologs and 68 by Veritas®. Cardiologs® had significantly fewer false positives than Veritas® (33 vs. 78). Accuracy of Cardiologs® (91.2%) was significantly higher than Veritas® (82.8%) ($p < 0.0001$), but not than Veritas® + physician (90.0%) ($p = 0.52$).

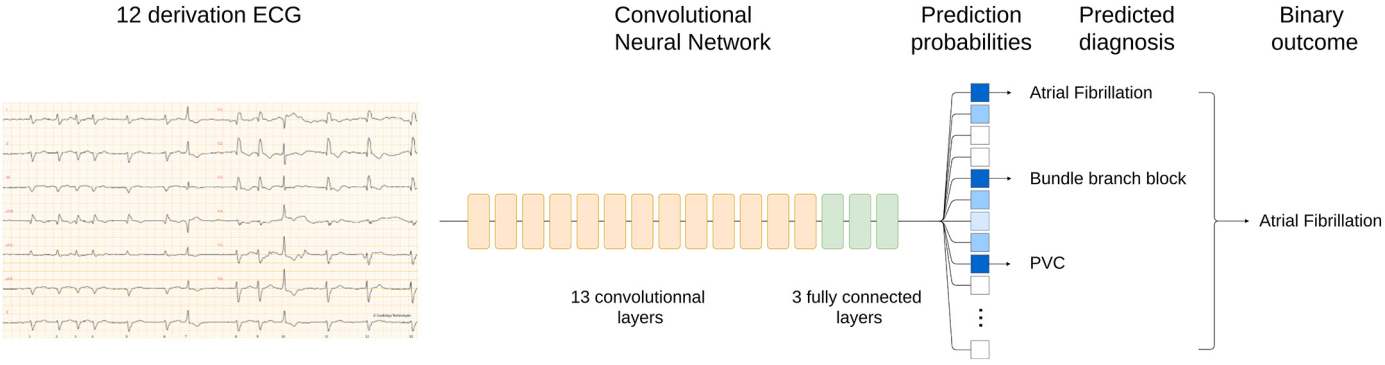

Fig. 1. Diagnostic prediction for resting ECG. A 12 lead ECG (left) is transformed by a convolutional neural network (center) into prediction probabilities for 76 different labels (high probabilities in blue, low ones in white). The predicted diagnosis is composed of all labels with probability higher than a given threshold (0.5 here). From this predicted diagnosis a binary outcome is calculated according to what is being studied (e.g. AF/no AF, AD/no AD).

Table 1
Final adjudicated rhythm out of 500 ECGs that were interpreted as AF by one or both algorithms.

	Atrial fibrillation	Atrial dysrhythmia but no atrial fibrillation	No atrial dysrhythmia	Total
Count	399 (79.80%)	15 (3.00%)	86 (17.20%)	500 (100.00%)
Associated ventricular rhythm				
Conducted to ventricle	388	13	77	478
Ventricular paced rhythm	9	2	0	11
Atrioventricular block with junctional escape rhythm	2	0	6	8
Atrioventricular block with ventricular escape rhythm	0	0	2	2
Ventricular tachycardia	0	0	1	1

The accuracy of the final physician overread of the Veritas® algorithm was 90%, with 28 false positives and 22 false negatives. When Veritas® was incorrect ($n = 99$), accuracy of Veritas® + physician was 61% (95% CI 52–71); in other words, 61% of Veritas® errors were corrected by the physician. Among those same ECGs, Cardiologs® accuracy was 90% (95% CI 82–94) ($p < 0.0001$), with both significantly higher sensitivity (100% vs. 48%, $P = 0.04$) and specificity (85% vs. 60%, $P < 0.0014$).

See Table 3 for the secondary outcome of AD vs. not. Again, Cardiologs had significantly more agreement and this was particularly striking in cases in which the Veritas® algorithm was incorrect.

The results were not different for ECGs with QRS duration > vs. < 120 ms, and for heart rates > vs. < 100 (data not shown).

Finally, for the analysis of data from the random sample of ED ECGs ($n = 1473$), see Table 4. There were 39 cases of AF, and 45 of AD. There was no significant difference in sensitivity of Cardiologs® vs. Veritas® for either AF (92% vs. 87%) or AD (93% vs. 93%). However, again, because of many more false positives by Veritas®, accuracy, specificity, and PPV were significantly higher for Cardiologs®.

4. Limitations

This study is retrospective, with many of the usual limitations. However, unlike many retrospective studies, all cases of Cardiologs®- or Veritas®-identified AF and AD were captured for the 6-month period, as all ECGs are kept in one database. Moreover, all automated Veritas® interpretations and all physician overreads of the Veritas® interpretation are recorded. Although cases of AF or AD that were not identified by one or both of the algorithms would have been missed, we analyzed a separate database to estimate sensitivity, which for AF was 93% for both algorithms. The reference diagnoses of AF and AD were based on ECG reading by experts, who are not infallible; a true reference standard, at least in difficult cases, would be the final clinical diagnosis, especially with an electrophysiologic study.

Furthermore, we are unable to compare Veritas® + physician to Cardiologs® + physician because only the Veritas® algorithm was used for clinical purposes and had a physician overread. It is likely that, since Cardiologs® was far more specific than Veritas® and even than Veritas® + physician, that Cardiologs® + physician would be more specific than Veritas® + physician. We only compared Cardiologs® to one particular conventional algorithm; it would have been ideal to compare the performance of Cardiologs®' algorithm to other existing algorithms. However, other algorithms are either embedded on recording devices not used (e.g., GE Marquette™ 12SL or Philips DXL), or are proprietary (like Glasgow® by Physio-Control) [9]. In both cases, we did not have access to these algorithms' interpretations.

5. Discussion

This study shows that the Cardiologs® DNN for 12-lead ECG interpretation has higher accuracy, specificity, and positive predictive value than both the Veritas® algorithm and the Veritas® algorithm with physician overread.

Atrial Fibrillation is a major world health problem, especially because AF is one of the most important etiologies of stroke. Accurate diagnosis of AF is imperative, as approximately 25–30% of ischemic stroke are associated with AF, and anticoagulation in AF patients is proven to prevent stroke [10,11]. The Global Burden of Disease Study by the World Bank estimated 2010 prevalence as 596 (men) and 373 (women) per 100,000, with almost 50 million people affected; the incidence since 1990 is increasing: disability has increased by 18% and mortality has doubled [12]. In 2016, the European estimates of 9 million patients with AF were predicted to increase to 17 million by 2030 with annual incidence increasing from 200,000 to 570,000 by 2030 [13].

12-Lead ECG instruments routinely provide automated computerized interpretations. There are several ways that healthcare providers (HCP) can use the automated interpretation. HCP may use it for efficiency and speed; the automated interpretation only needs to be confirmed or edited before being finalized, whereas a de novo interpretation would require more effort. HCP may use it to identify abnormalities they otherwise would not see: they may read the ECG first without looking at the automated interpretation, and only then look to see if there is something they might have missed. Or HCP may, in a hurry, accept an erroneous interpretation without scrutinizing the ECG. Finally, many HCP, including physicians, have become less skilled at ECG interpretation and may depend on the automated version.

Unfortunately, however, computer algorithms have mediocre performance [14]. When they are accurate, they have been shown to increase the accuracy of physician overread, but when inaccurate, they have been shown to lead physicians astray. This has been shown for cardiology fellows [15], for emergency physicians [16], and also for fully trained cardiologists, for whom the presence of an automated interpretation resulted in lower accuracy because automated errors were not corrected [17].

Table 2
Agreement of each method for the primary outcome: AF ($n = 399$) vs. Not-AF ($n = 101$). Total $n = 500$.

Comparison	Agree (n)	Accuracy, (Agreement) (95% CI)	Disagree (n)	*P* value
Expert vs. CardioLogs®	456	91.2% (88.4–93.4)	44 (33 FP, 11 FN)	Reference
Expert vs. Veritas®	401	80.2% (76.5–83.5)	99 (78 FP, 21 FN)	<0.0001
Expert vs. Veritas® + physician	450	90.0% (87.1–92.3)	50 (28 FP, 22 FN)	0.52
When Veritas® is incorrect, n = 99				
Expert vs. Veritas® + physician	61	62% (52–71)	38 (27 FP, 11 FN)	<0.0001
Expert vs. Cardiologs®	89	90% (82–94)	10 (10 FP, 0 FN)	

Reference = final expert interpretation.
FP = False positive, FN = False negative.

Table 3
Agreement of each method for the secondary outcome: AD ($n = 414$) vs. Not-AD ($n = 86$). Total n = 500.

Comparison	Agree	Accuracy, (Agreement) (95% CI)	Disagree	P value
Expert vs. CardioLogs®	475	95.0% (92.7–96.6)	25 (22 FP, 3 FN)	Reference
Expert vs. Veritas®	414	82.8% (79.2–85.9)	86 (74 FP, 12 FN)	<0.0001
Expert vs. Veritas® + physician	459	91.8% (89.1–93.9)	41 (27 FP, 14 FN)	0.04
When Veritas® is incorrect, n = 86				
Expert vs. Veritas® + physician	54	63% (52–72)	32 (26 FP, 6 FN)	<0.0001
Expert vs. Cardiologs®	76	86% (80–94)	10 (10 FP, 0 FN)	

Reference = final expert interpretation.
FP = False positive, FN = False negative.

Atrial dysrhythmias, and atrial fibrillation in particular, are frequently misdiagnosed by computer algorithms and then by the physician who overreads them. Shah and Rubin studied the computer rhythm interpretation of 2160 12-lead ECGs, compared to 2 cardiologists [18]. Among the 2112 that the computer could interpret, the sensitivity and specificity were, respectively, 95% and 66.3% for sinus rhythm and 71% and 95% for atrial fibrillation. Taggar et al. noted computer over-interpretation (false positives) of AF in 9 to 19% of ECGs [19]. Hwan Bae et al. found computer under-diagnosis on 9.3% of 840 AF ECGs, and over-diagnosis in 11.3%, with 7.8% of misdiagnoses uncorrected by the physician [20]. Among, 2447 ECGs, Mant et al. found 83.3% sensitivity and 99% specificity of the Biolog interpretive software; however, the reference experts only disagreed in 7 ECGs (0.27%), which suggests that only uncomplicated ECGs were included [21]. Poon et al. found 90.8% sensitivity and 98.9% specificity of the General Electric (GE) algorithm for AF, but they excluded cases with ventricular paced rhythm (VPR), in which the underlying atrial rhythm was frequently AF; we did not exclude VPR, as accurate identification of AF in VPR is important for determination of stroke risk [22]. Thus, both overdiagnosis and underdiagnosis of AF are well-recognized problems in computer-interpreted electrocardiograms, and they adversely affect patient management with either inappropriate anticoagulation and/or antidysrhythmics or undertreatment [1,14]. In our study, even physicians' overread of the Veritas® algorithm disagreed with the reference interpretation in 38 cases (7.6% of all cases, and 62% of Veritas® misdiagnoses).

Thus, the lower performance of the Veritas + physician compared to the DNN algorithm is due to the excellent performance of the DNN, but also due to the bias that an incorrect automated diagnosis introduces into the final diagnosis [1]. Misclassification by the DNN algorithm could be due to the under-representation of certain types of samples within the training set, such as providing it with combinations of AF with other pathologies, including myocardial infarction, ventricular hypertrophy, and others. We analyzed the influence of HR and QRS duration on the performance of the algorithm without finding a correlation, as reported in the results section.

AF burden, defined as the amount of time spent in AF, is directly associated with incidence of stroke [23]. Ambulatory monitoring detects far more atrial fibrillation in patients with cryptogenic stroke than conventional follow-up [24]. Thus, detection of AF, without false alarms, is an important goal of automated ECG interpretation. In a previous study of Cardiologs® DNN in interpretation of the MIT-BIH arrhythmia public database of 30-minute Holter recordings [25], Cardiologs® was far more specific (98.5% vs. 82.8%) and accurate (98.3% vs. 85.5%) at temporally localizing atrial fibrillation than previously published methods, most of which rely on algorithmic instructions regarding the RR interval [26]. In a recent study comparing a different DNN to a cardiology committee consensus in single lead diagnosis of 12 distinct dysrhythmias, at a specificity of 94.1% for AF, the DNN sensitivity was 86.1%, vs. 71.0% for cardiologists [27].

Cardiologs®' algorithm predicts diagnostics from raw 12-lead ECG electrode recordings; it outputs the probability of presence of 76 different labels which can correspond both to general classes of pathologies (e.g. atrial dysrhythmia) or to specific pathologies within these classes (e.g. atrial fibrillation). Labels are non-exclusive, meaning that if both the general and specific label have a probability >0.5, they will appear in the output. The sensitivity of the algorithm can be augmented by lowering this detection threshold, though this will decrease the specificity. Cardiologs® uses a single model to predict the presence or absence of all labels simultaneously. This enables the model to take into account the dependence between pathologies. For example, when a patient has a pre-excitation syndrome, the repolarization is affected by this pre-excitation. The algorithm can then learn to disregard the ST segment abnormality of a pseudoinfarction pattern in order to avoid the false alarm of ischemia.

The Cardiologs® DNN is a convolutional neural network (CNN) and was detailed in our previous paper [8]. In order to avoid confusion for the reader regarding CNN vs. DNN, CNN is a specific type of DNN; that is, all CNN are DNN, but not the converse [3,28]. The neural network was implemented using the Keras framework with a Tensorflow backend (Google; Mountain View, CA, USA). Model training and testing were performed using a K-80 (NVIDIA) graphics processing unit. Importantly, the model was trained using data from multiple ECG recording machines; only a small portion of training examples were recorded using Mortara® ECG machine (5000 out of 80,000), which is the machine used in this study. Therefore, the algorithm is device independent. On the other hand, the Veritas® algorithm is embedded on the Mortara® ECG machine and thus might be expected to be most accurate on this device.

Automated interpretations have until now mostly been conventional instructional algorithms of "if, then." [29] Some algorithms use advanced machine learning techniques for finding individual abnormalities such as atrial fibrillation, but are limited to one abnormality at a time, whereas the Cardiologs® algorithm uses DNN technology to

Table 4
Diagnostic performance for AF vs. no-AF and AD vs. no-AD in the study of 1473 random ECGs [8].

	Atrial fibrillation ($n = 39$)			Atrial dysrhythmias (n = 45)		
	Cardiologs®	Veritas®	p value			p value
Sensitivity	92	87	0.46	93	93	1.0
Specificity	99.5	98.7	0.03	99.3	97.6	0.0003
Accuracy	99.3 (98.8–99.6)	98.4 (97.7–99.0)	0.023	99.1 (98.5–99.5)	97.5 (96.6–98.2)	0.0006
Positive predictive value	84% (36/43)	65% (34/52)	0.04	81% (42/52)	55% (42/76)	0.003
Negative predictive value	99.8%	99.6%	NS	99.8%	99.8%	1.0
Agree with reference	1463	1450		1460	1436	
Disagree with reference	10	23		13	37	

learn from, and interpret, the whole 12-lead ECG. Deep neural networking machine learning artificial intelligence does not rely on rules and can handle multiple abnormalities at once. Cardiologs® DNN uses large amounts of data to "train" the computer by labeling each case according to an unlimited number of pre-specified abnormalities, so that the algorithm may discern what characteristics of ECGs are associated with any given abnormality. Unlike expert humans, who may retire or die, the DNN keeps learning and improving indefinitely as more and more data, with accurate corresponding expert interpretations and outcomes, accrue.

Previous neural networks for 12-lead ECG interpretation also had many limitations. Though most can learn on their own, they are generally trained to find ECG abnormalities based on input of a simplified representation of an ECG, with manually written algorithms that assess such features as the ST elevation value [30,31]. Some have a single simplified output such as acute myocardial infarction [30,31] or atrioventricular block [32]. Some have a list of exclusive outputs defining an ECG or a beat [such as left bundle branch block vs. right bundle branch block vs. paced rhythm vs. ventricular vs. normal vs. other] [33–35].

DNN have previously been shown to be accurate for isolated aspects of the ECG, such as rhythm diagnosis, for which it exceeded cardiologists' performance for arrhythmia detection [36]. Cardiologs® Technologies' new DNN algorithm, which is not limited to interpretation of any one ECG label, can interpret the entire 12-lead ECG, including rhythm and QRS-T-U waves.

Previously, we showed that the Cardiologs® algorithm had more accurate global 12-lead ECG interpretation (for both rhythm and QRSTU abnormalities) than Veritas®, with a higher accuracy and positive predictive value in finding major ECG abnormalities and also in overall interpretation [8]. In that study of 1473 randomly selected ED cases, there were 45 cases of AD, 39 of which were AF, and from this database we show equal sensitivity, but better specificity and accuracy of Cardiologs® vs. Veritas®. In order to more accurately estimate the accuracy and specificity of AF and AD in particular, we expanded the cohort by randomly choosing 500 out of all 858 during the same time period that were identified by one or the other algorithms as AD. This allowed us to better demonstrate the accuracy, specificity and positive predictive value of Cardiologs® for AF and AD, and, importantly, to compare it to the physician overread of the ECG.

6. Conclusion

In the diagnosis of atrial fibrillation, Cardiologs® new convolutional deep neural network was significantly more accurate and specific than the Veritas® conventional algorithm, and equally sensitive. Among those misdiagnosed by Veritas®, it was more specific and accurate than even the physician overread. It was also more specific and accurate in diagnosing any atrial dysrhythmia.

Funding

All funding was by Cardiologs® technologies.

The first author, Dr. Stephen W. Smith, received no funding and received no financial support of any kind from any source.

Declaration of competing interest

Stephen W. Smith – None.

Christophe Gardella – employed by Cardiologs®, shareholder in Cardiologs®.

Jeremy Rapin – formerly employed by Cardiologs®. Shareholder in Cardiologs®.

Jia Li – employed by Cardiologs®, shareholder in Cardiologs®.

Yann Fleureau – CEO of Cardiologs®, shareholder.

Brooks Walsh – In a different study by Cardiologs®, paid a set fee for blind interpretation of ECGs.

William Fennell – None.

Arnaud Rosier – shareholder in Cardiologs®.

Laurent Fiorina – employed by Cardiologs®.

References

[1] F. Bogun, D. Anh, G. Kalahasty, et al., Misdiagnosis of atrial fibrillation and its clinical consequences, Am. J. Med. 117 (9) (2004) 636–642 , Internet. Available from: https://doi.org/10.1016/j.amjmed.2004.06.024.

[2] P. Tschandl, C. Rosendahl, B.N. Akay, et al., Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks, JAMA Dermatol. 155 (1) (2019) 58–65, https://doi.org/10.1001/jamadermatol.2018.4378.

[3] A. Esteva, B. Kuprel, R.A. Novoa, et al., Dermatologist-level classification of skin cancer with deep neural networks, Nat. Res. 542 (2017) 115–118.

[4] A. Lee, P. Taylor, J. Kalpathy-Cramer, A. Tufail, Machine learning has arrived! Ophthalmology 124 (12) (2017) 1726–1728.

[5] S. Chilamkurthy, R. Ghosh, S. Tanamala, et al., Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study, Lancet 392 (10162) (2018) 2388–2396 , Internet. Available from https://www.ncbi.nlm.nih.gov/pubmed/30318264.

[6] P. Rajpurkar, J. Irvin, K. Zhu, et al., CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays With Deep Learning, Available from: doi.org/arXiv:1711.05225v3 2017.

[7] D. Ardila, A.P. Kiraly, S. Bharadwaj, et al., End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography, Nat. Med. (2019)https://doi.org/10.1038/s41591-019-0447-x Internet. Available from:.

[8] S.W. Smith, B. Walsh, K. Grauer, et al., A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation, J. Electrocardiol. 52 (2019) 88–95 , [Internet]. Available from: https://doi.org/10.1016/j.jelectrocard.2018.11.013.

[9] Physio-Control, Inc., Statement of Validation and Accuracy for the Glasgow® 12-Lead ECG Analysis Program Version 27 [Internet], Available from: https://www.physio-control.com/uploadedFiles/learning/clinical-topics/Glasgow%2012-lead%20ECG%20Analysis%20Program%20Statement%20of%20Validation%20and%20Accuracy%203302436.A.pdf 2009.

[10] C. Marini, F. De Santis, S. Sacco, et al., Contribution of atrial fibrillation to incidence and outcome of ischemic stroke: results from a population-based study, Stroke 36 (6) (2005) 1115–1119 , Internet. Available from https://www.ncbi.nlm.nih.gov/pubmed/15879330.

[11] N. Hannon, O. Sheehan, L. Kelly, et al., Stroke associated with atrial fibrillation – incidence and early outcomes in the North Dublin population stroke study, Cerebrovasc. Dis. 29 (2010) 43–49, https://doi.org/10.1159/000255973 Cerebrovasc Dis 29(1):43–9.

[12] S.S. Chugh, R. Havmoeller, K. Narayanan, et al., Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study, Circulation 129 (8) (2014) 837–847 , [Internet]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24345399.

[13] M. Zoni-Berisso, F. Lercari, T. Carazza, S. Domenicucci, Epidemiology of atrial fibrillation: European perspective, Clin. Epidemiol. 6 (2014) 213–220 , Internet. Available from https://www.ncbi.nlm.nih.gov/pubmed/24966695.

[14] J. Schlapfer, H.J. Wellens, Computer-interpreted electrocardiograms: benefits and limitations, J. Am. Coll. Cardiol. 70 (9) (2017) 1183–1192.

[15] T. Novotny, R. Bond, I. Andrsova, et al., The role of computerized diagnostic proposals in the interpretation of the 12-lead electrocardiogram by cardiology and non-cardiology fellows, Int. J. Med. Inform. 101 (5) (2017) 85–92.

[16] P. Martinez-Losas, J. Higueras, J.C. Gomez-Polo, et al., The influence of computerized interpretation of an electrocardiogram reading, Am. J. Emerg. Med. 34 (10) (2016) 2031–2032 , Internet. Available from: https://doi.org/10.1016/j.ajem.2016.07.029.

[17] D. Anh, S. Krishnan, F. Bogun, Accuracy of electrocardiogram interpretation by cardiologists in the setting of incorrect computer analysis, J. Electrocardiol. 39 (3) (2006) 343–345 , [Internet]. Available from:. https://doi.org/10.1016/j.jelectrocard.2006.02.002.

[18] A.P. Shah, S.A. Rubin, Errors in the computerized electrocardiogram interpretation of cardiac rhythm, J. Electrocardiol. 40 (5) (2007) 385–390 , Internet. Available from https://www.ncbi.nlm.nih.gov/pubmed/17531257.

[19] J.S. Taggar, T. Coleman, S. Lewis, C. Heneghan, M. Jonesa, Accuracy of methods for diagnosing atrial fibrillation using 12-lead ECG: a systematic review and meta-analysis, Int. J. Cardiol. 184 (2015) 175.

[20] M. Hwan Bae, J. Hoon Lee, D. Heon Yang, et al., Erroneous computer electrocardiogram interpretation of atrial fibrillation and its clinical consequences, 35 (6) (2012) 348–353.

[21] J. Mant, D.A. Fitzmaurice, F.D. Hobbs, et al., Accuracy of diagnosing atrial fibrillation on electrocardiogram by primary care practitioners and interpretative diagnostic software: analysis of data from screening for atrial fibrillation in the elderly (SAFE) trial, BMJ 335 (7616) (2007) 380 , Internet. Available from https://www.ncbi.nlm.nih.gov/pubmed/17604299.

[22] K. Poon, P.M. Okin, P. Kligfield, Diagnostic performance of a computer-based ECG rhythm algorithm, J. Electrocardiol. 38 (3) (2005) 235–238 , Internet. Available from https://www.ncbi.nlm.nih.gov/pubmed/16003708.

[23] L.Y. Chen, M.K. Chung, L.A. Allen, et al., Atrial fibrillation burden: moving beyond atrial fibrillation as a binary entity: a scientific statement from the American Heart Association, Circulation 137 (20) (2018) e623–e644 , Internet. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29661944.

[24] T. Sanna, H.-C. Diener, R.S. Passman, et al., Cryptogenic stroke and underlying atrial fibrillation, 370 (2014) 2478–2486.
[25] P. Schluter, Performance measures for arrhythmia detectors, Comput. Cardiol. 7 (1980).
[26] J. Li, J. Rapin, A. Roser, S.W. Smith, Y. Fleureau, P. Taboulet, Deep neural networks improve atrial fibrillation detection in Holter: first results, Eur. J. Prev. Cardiol. 23 (2016) 41 2_Suppl.
[27] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, et al., Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, Nat. Med. 25 (1) (2019) 65–69 , Internet. Available from: https://doi.org/10.1038/s41591-018-0268-3.
[28] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444 , Internet. Available from: https://doi.org/10.1038/nature14539.
[29] J. Prineas, R.S. Crow, H. Blackburn, The Minnesota Code Manual of Electrocardiographic Findings: Standards and Procedures for Measurement and Classification, Springer, Littleton, Mass, 2009.
[30] B. Heden, H. Ohlin, R. Rittner, L. Edenbrandt, Acute myocardial infarction detected in the 12-lead ECG by artificial neural networks, Circulation 96 (1997) 1798–1802.
[31] J. Kojuri, R. Boostani, P. Dehghani, F. Nowroozipour, N. Saki, Prediction of acute myocardial infarction with artificial neural networks in patients with nondiagnostic electrocardiogram, J. Cardiovasc. Dis. Res. 6 (2) (2015) 51–59.
[32] S. Meghriche, A. Draa, M. Boulemden, On the analysis of a compound neural network for detecting AtrioVentricular heart block (AVB) in an ECG signal, Int. J. Biol. Life Sci. 4 (1) (2008) 1–11.
[33] G. Bortolan, R. Degani, J.L. Willems, Neural Networks for ECG Classification, 1990.
[34] H. Rai, A. Trivedi, S. Shukla, V. Dubey, ECG Arrhythmia Classification Using Daubechies Wavelet and Radial Basis Function Neural Network, 2012.
[35] S. Kiranyz, T. Ince, R. Hamila, M. Gabboui, Convolutional Neural Networks for Patient-Specific ECG Classification, 2015.
[36] P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, A.Y. Ng, Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks, Available from: https://stanfordmlgroup.github.io/projects/ecg/ 2017.