An artificial intelligence–enabled ECG algorithm for comprehensive ECG interpretation: Can it pass the 'Turing test'?

Anthony H. Kashou, MD,* Siva K. Mulpuru, MD, FHRS,†
Abhishek J. Deshmukh, MBBS, FHRS,† Wei-Yin Ko, MS,† Zachi I. Attia, PhD,†
Rickey E. Carter, PhD,‡ Paul A. Friedman, MD, FHRS,† Peter A. Noseworthy, MD, FHRS†

*From the *Department of Medicine, Mayo Clinic, Rochester, Minnesota, †Department of Cardiovascular Medicine, Mayo Clinic, Rochester, Minnesota, and ‡Department of Health Sciences Research, Mayo Clinic, Jacksonville, Florida.*

OBJECTIVE To develop an artificial intelligence (AI)–enabled electrocardiogram (ECG) algorithm capable of comprehensive, human-like ECG interpretation and compare its diagnostic performance against conventional ECG interpretation methods.

METHODS We developed a novel AI-enabled ECG (AI-ECG) algorithm capable of complete 12-lead ECG interpretation. It was trained on nearly 2.5 million standard 12-lead ECGs from over 720,000 adult patients obtained at the Mayo Clinic ECG laboratory between 2007 and 2017. We then compared the need for human over-reading edits of the reports generated by the Marquette 12SL automated computer program, AI-ECG algorithm, and final clinical interpretations on 500 randomly selected ECGs from 500 patients. In a blinded fashion, 3 cardiac electrophysiologists adjudicated each interpretation as (1) ideal (ie, *no* changes needed), (2) acceptable (ie, *minor* edits needed), or (3) unacceptable (ie, *major* edits needed).

RESULTS Cardiologists determined that on average 202 (13.5%), 123 (8.2%), and 90 (6.0%) of the interpretations required major edits from the computer program, AI-ECG algorithm, and final clinical interpretations, respectively. They considered 958 (63.9%), 1058 (70.5%), and 1118 (74.5%) interpretations as *ideal* from the computer program, AI-ECG algorithm, and final clinical interpretations, respectively. They considered 340 (22.7%), 319 (21.3%), and 292 (19.5%) interpretations as *acceptable* from the computer program, AI-ECG algorithm, and final clinical interpretations, respectively.

CONCLUSION An AI-ECG algorithm outperforms an existing standard automated computer program and better approximates expert over-read for comprehensive 12-lead ECG interpretation.

KEYWORDS Artificial intelligence; Convolutional neural network; ECG; ECG interpretation; Electrocardiogram; Electrocardiography

(Cardiovascular Digital Health Journal 2021;2:164–170) © 2021 Heart Rhythm Society. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Nearly 1 century after Willem Einthoven was awarded the Nobel Prize for demonstrating the electrocardiogram (ECG) could record cardiac biosignals, the ECG continues to serve as an accessible, inexpensive, and noninvasive means to assess cardiac activity and function. Since the 1950s, the development of an analog-to-digital converter that enabled processing of digital information[1] has allowed computerized ECG interpretation algorithms to automatically extract, analyze, and interpret ECGs[2,3] in order to minimize interpretation error, expedite clinical decision making, and optimize workflow.[4] These programs have since become a routine part of ECG interpretation in clinical practice. However, these technologies remain inherently prone to error.[4–8] Furthermore, their

influence on the over-reading provider's final interpretation is also profound,[9–12] which carries the risk of perpetuating incorrect interpretations and consequent patient harm.[8]

Simultaneous advances in computing power and digitized data availability has catalyzed the application of artificial intelligence (AI) to the ECG. While AI-enabled ECG (AI-ECG) algorithms have demonstrated the ability to recognize individual ECG patterns and diagnoses, they have been limited to single-lead ECGs or in their diagnostic scope.[13–15] There still remains no model capable of generating a comprehensive but parsimonious human-like 12-lead ECG interpretation.

In this study, we developed a novel AI-ECG algorithm that uses a convolutional neural network (CNN) for feature extraction and a transformer network to translate the ECG features into a text sequence that follows the typical language used by human ECG interpretations, thereby creating an ECG interpretation that mirrors natural language. We then

Address reprint requests and correspondence: Dr Anthony H. Kashou, Department of Medicine, Mayo Clinic, 200 First St SW, Rochester, MN 55905. E-mail address: kashou.anthony@mayo.edu.

compared the need for human over-read edits in the interpretations generated by the Marquette 12SL automated computer program, a novel AI-ECG algorithm, and clinical interpretation processes from 500 randomly selected 12-lead ECGs.

Methods

Data from the Mayo Clinic digital vault was only used from individuals who provided consent to use of their anonymized records for research. The authors are unable to make the data publicly available, as it originates from Mayo Clinic's ECG database, which has patient identifying information. However, the authors have made the methods available to the reader.

AI-ECG algorithm development, training, validation, and testing

Details of the model development, training, validation, and testing were previously described.[16] Briefly, we developed an AI-ECG algorithm capable of complete 12-lead ECG interpretation. Model development was inspired by the inverse cooking system described by Salvador and colleagues.[17] This ECG transformer model uses a CNN for ECG feature extraction and a transformer network for translating ECG features into ECG codes. The model was trained to identify 66 discrete ECG diagnostic codes, including primary and secondary rhythms, axis deviation, chamber enlargement, atrioventricular and intraventricular conduction delay, myocardial ischemia, waveform abnormalities, clinical disorders, and pacemakers. Since each ECG can have multiple codes, this was considered a multilabel task. For each ECG, the model created a binary evaluation of whether the code was present. The network architecture contained 11 bottleneck ResNet blocks, which were made up of 33 convolutional layers. It had 3 dimensions, and the convolutional kernels were either 3×3 or 1×1 in size (ie, the bottlenecks) with a stride of 2 for each block. The final block had a channel output of 2048. The CNN condensed the 10-second 12-lead ECG at 500 Hz, with a size of 5000×12, into a 2048×3 matrix of relevant features. The matrix was then reshaped into 12×512. The 512 represents the embedding size in order for each ECG to be used as a sequence of 12 embedding for the transformer. In order to encourage generalization of features, an additional set of 1000 embeddings (ECG concepts; selected after trialing different numbers of ECG concepts) were used for the transformer encoder. The embeddings chosen were enough to map the complexities of the features, while not being too large such that there would be too many similar concepts. Given that each of the 12 embeddings for every ECG is likely unique, this creates a large, information-sparse corpus. Mapping each embedding to these reusable abstract concepts forces the model to learn the relationship of the unique embeddings, as the model must optimize the generalizability of the concept embedding being mapped to. The result is then concatenated with the original 12-embedding sequence of the ECG to form a 24-embedding sequence (24×512)

that is the input for the transformer decoder. This matrix reshaping from 3×2048 to 12×512 allowed for a substantial amount of embeddings (1×512) in a sequence (12) for the transformer model to learn the association.

The transformer decoder is used to translate the combination of ECG and ECG concepts into a string of ECG codes that approximates the text of a human ECG interpretation. Given that ECG code tokens such as "possible" or "cannot rule out" cannot be derived from the ECG signal, the self-attention mechanism[18] of the transformer model plays a significant role in learning the context where these tokens might appear in the ECG sequence. For sequence generation, we simply used the ECG codes with the maximum probability of appearing given the sequence to select the next token. The training process used teacher forcing for the first 2 epochs, then all predicted sequences were as is. We used AdamW as the training optimizer and Cross Entropy as the loss function. The model was created using the PyTorch deep learning library.[19]

The network was developed on nearly 2.5 million standard 12-lead ECGs from over 720,000 adult patients at the Mayo Clinic. All data used for initial AI-ECG algorithm development were obtained electronically without manual ECG review. The CNN was pretrained on identifying the visible features in the ECG codes via multilabel classification. The CNN uses the raw ECG data tracings as input and outputs discrete diagnostic ECG labels. Each of the 66 discrete, structured labels was represented by one-hot encoding to account, since there can be multiple visible features in each ECG.

Pretraining the CNN model reduced the amount of time required to train the entire CNN-transformer model owing to its large network size. The last fully connected layer for classification was then discarded, leaving only the convolutional layers for extracting ECG features. The transformer itself encoded the extracted ECG feature embeddings and was then decoded back into ECG code embeddings. The word tokens in the ECG codes were used along with the codes for the visible features, which usually forms a short sentence-like sequence for each ECG to train the transformer in a supervised manner.

The data sets for model derivation included the following: (1) a training set of 1,749,654 ECGs, (2) a validation set composed of 249,951 ECGs, and (3) a testing set composed of 499,917 ECGs. For training, the learning rate was started at 0.0001. The learning rate was then manually decreased over time. The model was trained for as long as necessary to clear overfitting of the training set. Validation loss was monitored, and the training process was halted if the validated loss stopped decreasing for 10 epochs. Using this approach, we were able to select the best-performing model checkpoint on the validation set and apply this to the final testing set. We selected the optimal network by convoluting between the leads rather than treating the 12 leads independently. This mimicked human interpreters' approach by evaluating the composite ECG based on the relative relationships of the raw voltage values. Final performances for each ECG are presented in Tables 2–6 of previously published work.[16]
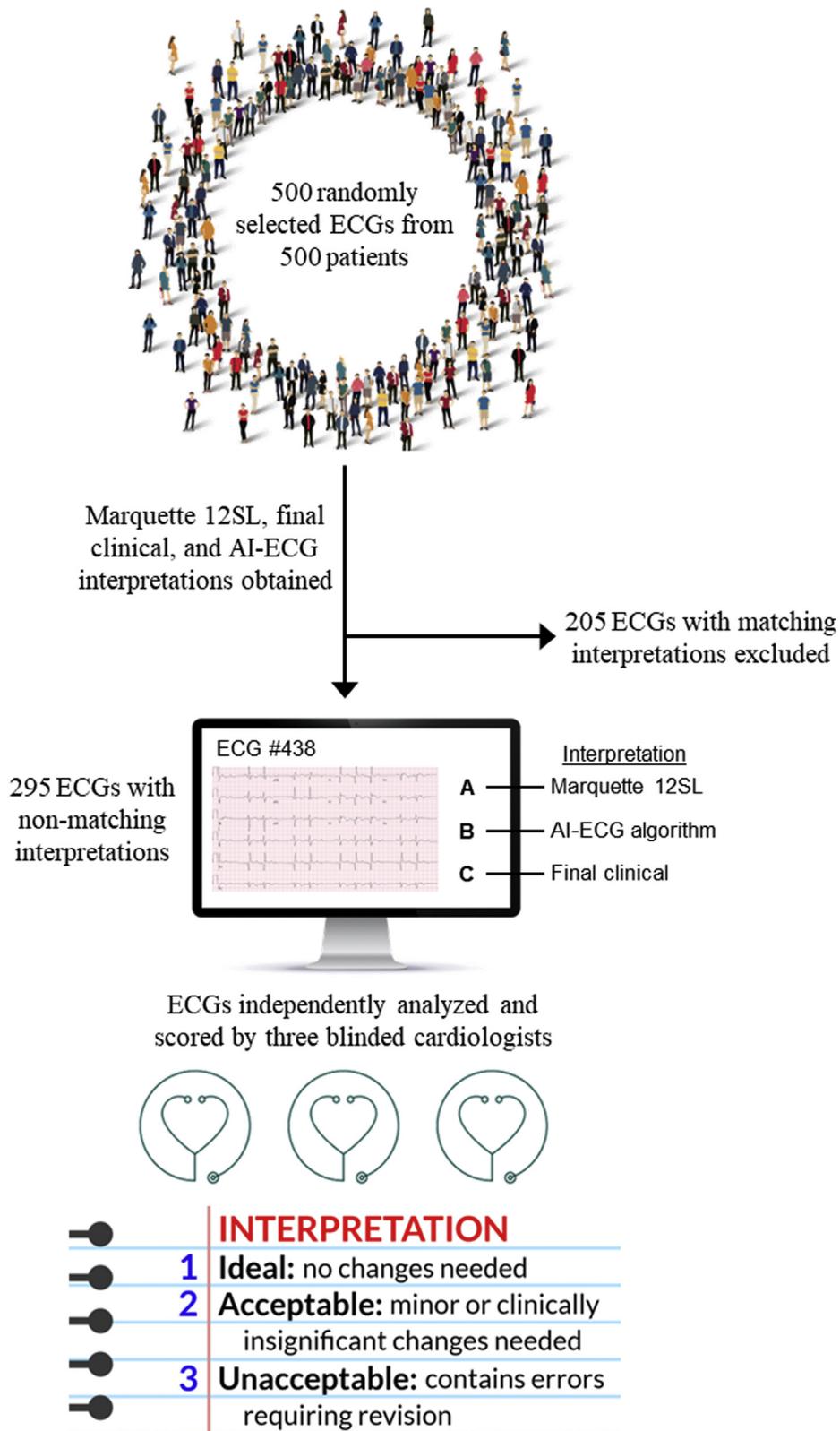
Figure 1 Study design. AI-ECG = artificial intelligence–enabled electrocardiogram; ECG = electrocardiogram.

Table 1 Diagnostic performance and comparisons of each interpretation method

Interpretation	Interpreter 1	Interpreter 2	Interpreter 3	Average
Computer-generated				
Ideal	62.0%	65.0%	64.6%	63.9%
Acceptable	23.6%	18.4%	26.0%	22.7%
Unacceptable	14.4%	16.6%	9.4%	13.5%
AI-ECG				
Ideal	70.2%	66.6%	74.8%	70.5%
Acceptable	22.8%	21.2%	19.8%	21.3%
Unacceptable	7.0%	12.2%	5.4%	8.2%
Final clinical				
Ideal	76.8%	65.2%	81.4%	74.5%
Acceptable	18.6%	24.2%	15.8%	19.5%
Unacceptable	4.6%	10.6%	2.8%	6.0%

The performance of the computer-generated, AI-ECG, and final clinical interpretations from each participating cardiologist and their average scores are displayed. Interpretation scoring system: *ideal* indicates no changes needed to the interpretation; *acceptable* indicates minor or clinically insignificant changes needed to the interpretation; and *unacceptable* indicates that the interpretation contains errors requiring revision.

AI-ECG = artificial intelligence–enabled electrocardiogram.

ECG selection and interpretation collection for comparison

We randomly selected 500 standard 10-second, 12-lead ECGs from 500 patients ≥18 years of age obtained at the Mayo Clinic ECG laboratory between 2007 and 2017 that were not used in the model development stage. ECGs were recorded at a sampling rate of 500 Hz using the GE-Marquette 12SL ECG analysis program (GE Healthcare, Milwaukee, WI) and the raw data were stored using the MUSE data management system (GE Healthcare). While ECGs were not reviewed for artifact or lead misplacement prior to incorporation into the study, they were all previously reviewed at the time of ECG recording and considered clinically acceptable. No patient or ECG was excluded from the study. The Marquette 12SL automated computer program and final clinical interpretations were retrieved. Every final clinical interpretation obtained was completed by a board-certified, practicing cardiologist with access to the patient's clinical information at the time of the ECG recording. The digital data were fed into the developed AI-ECG algorithm, which yielded an independent and comprehensive interpretation that was collected and used for comparison. None of the ECGs included in the head-to-head diagnostic analysis were used in training the AI-ECG algorithm. In the end, all ECGs used for final assessment had a Marquette 12SL automated computer–generated, final clinical, and AI-ECG algorithm–generated interpretation.

Cardiac electrophysiologist evaluation

All ECGs and corresponding interpretations (ie, Marquette 12SL automated computer–generated, final clinical, and AI-ECG algorithm–generated) were compiled. Of the randomly selected 500 ECGs, 205 ECGs did not require cardiac electrophysiologist review because all methods generated identical interpretations (Figure 1). The remaining 295 ECGs with non-matching (discordant) interpretations were used for cardiac electrophysiologist review and analysis. Interpretations were presented to the readers in a randomized and unlabeled format without any corresponding clinical information.

Three blinded board-certified, practicing, and experienced cardiac electrophysiologists (expert over-reading cardiologists) independently analyzed the 295 ECGs and their corresponding interpretations (ie, Marquette 12SL automated computer–generated, final clinical, and AI-ECG algorithm–generated interpretations for each ECG). None of the expert over-reading cardiologists were involved with compiling the ECGs and interpretations, and therefore they remained blinded throughout the entire process. The expert over-reading cardiologists were not provided any clinical data (eg, age, sex, medical history, previous ECG, etc) or any identifying patient information when presented with each ECG. Expert over-reading cardiologists only had access to a document with the 295 standard 10-second, 12-lead ECGs and the corresponding unlabeled, randomized, non-matching computer-generated, final clinical, and AI-ECG algorithm–generated interpretations.

Expert over-reading cardiologists were asked to examine each ECG and its corresponding interpretations, and to give an accuracy score for each interpretation:

- 3 = unacceptable interpretation (ie, the interpretation contains errors requiring revision)
  - o Example: ECG demonstrating right bundle branch block, but is interpreted as "bundle branch block." This would be considered not specific enough and receive a score of 3.
- 2 = acceptable interpretation (ie, only minor or clinically insignificant changes to the interpretation would be needed)
  - o Example: ECG demonstrating clear left atrial enlargement, but is interpreted as "borderline left atrial enlargement." This would be considered clinically acceptable with only minor changes and receive a score of 2.
- 1 = ideal interpretation (ie, no additional changes to the interpretation would be needed)

Statistical analysis

The primary outcome was the diagnostic accuracy of each ECG interpretation method. Diagnostic accuracy was based

Table 2 Interpreter agreement for each interpretation method assessed individually (κ) and collectively (α).

Interpretation	I₁ vs I₂	I₁ vs I₃	I₂ vs I₃	All interpreters
Computer-generated	κ = 0.550	κ = 0.497	κ = 0.441	α = 0.497
AI-ECG	κ = 0.377	κ = 0.382	κ = 0.356	α = 0.367
Final clinical	κ = 0.188	κ = 0.244	κ = 0.271	α = 0.224

α = Krippendorff's alpha coefficient; AI-ECG = artificial intelligence–enabled electrocardiogram; I = interpreter; κ = kappa coefficient.

off the accuracy scores provided by the expert over-reading cardiologists. The diagnostic performance and mean and median interpreter composite scores for each interpretation method were compared using unadjusted and Bonferroni-adjusted P values. The secondary outcome was the percentage of interpretations considered clinically acceptable by each method. Interpretations considered acceptable to clinical standards were those that received a score of 1 or 2. Inter-interpreter variability was also assessed between individual interpreters using the kappa coefficient and collectively using Krippendorff's alpha coefficient.[20,21]

Results

Of the 500 unselected ECGs, 205 ECGs were perfect interpretation matches and were excluded from interpreter evaluation. These 205 matching ECG interpretations were considered ideal interpretations (score of 1) and were incorporated in the final results as such. The remaining 295 ECGs and their corresponding interpretations were used to assess the need for additional edits by the interpreters.

Interpretation accuracy

Cardiac electrophysiologist assessment of interpretations obtained from each interpretation method and their summative averages are displayed in Table 1. Expert over-reading cardiologists determined that 202 (13.5%), 123 (8.2%), and 90 (6.0%) of the interpretations required edits (*unacceptable*) in the Marquette 12SL automated computer program, AI-ECG algorithm, and final clinical interpretations, respectively. The expert over-reading cardiologists considered 958 (63.9%), 1058 (70.5%), and 1118 (74.5%) interpretations as *ideal* in the Marquette 12SL automated computer program, AI-ECG algorithm, and final clinical interpretations, respectively. They considered 340 (22.7%), 319 (21.3%), and 292 (19.5%) interpretations as *acceptable* the Marquette 12SL automated computer program, AI-ECG algorithm, and final clinical interpretations, respectively.

Mean (standard deviation) and median (quartile 1, quartile 3) interpreter composite scores for the Marquette 12SL automated computer, AI-ECG algorithm, and final clinical interpretations were 1.497 (0.631) and 1.167 (1.000, 2.000), 1.377 (0.441) and 1.000 (1.000, 1.667), and 1.315 (0.527) and 1.000 (1.000, 1.667), respectively. Unadjusted and Bonferroni-adjusted P values for interpreter composite scores of the Marquette 12SL automated computer and AI-ECG algorithm interpretations were <.0001 and .0001, respectively. Unadjusted and Bonferroni-adjusted P values for interpreter composite scores of the AI-ECG algorithm and final clinical interpretations were .0250 and .0740, respectively. Unadjusted and Bonferroni-adjusted P values for interpreter composite scores of the Marquette 12SL automated computer and final clinical interpretations were both <.0001.

Among the 3 interpretation methods, the expert over-reading cardiologists considered 86.6%, 91.8%, and 94.0% of the Marquette 12SL automated computer, AI-ECG

algorithm, and final clinical interpretations as ideal (score of 1) or clinically acceptable (score of 2), respectively. When combining ideal (score of 1) and acceptable (score of 2) interpretations (ie, clinically acceptable interpretations), Krippendorff's alpha coefficients of agreement among all were 0.497, 0.367, and 0.224 for the Marquette 12SL automated computer, AI-ECG algorithm, and final clinical interpretations, respectively. Kappa coefficients for inter-rater agreement are reported in Table 2.

Discussion

This is the first study to assess and directly compare the need for additional human over-reading edits for the Marquette 12SL automated computer–generated, AI-ECG algorithm–generated, and final clinical interpretations. Our analysis demonstrates that an AI-ECG algorithm outperforms an existing standard automated computer program and better approximates expert cardiologist over-read for comprehensive standard 12-lead ECG interpretation. These data suggest that an AI-ECG algorithm may serve as an alternative, and perhaps more accurate, means to provide an initial interpretation for clinicians compared to conventional computer algorithms.

Clinical value

Standard computer-generated ECG interpretations provide several benefits to clinical practice. They can improve interpretation efficiency and expedite patient care.[4] They can also alert providers to abnormalities that may otherwise go overlooked. Furthermore, medical providers lacking confidence in their ECG interpretation skills may rely on their interpretation to direct patient care. Multiple studies have shown that correct computer-generated interpretations can improve physician over-read accuracy; however, incorrect automated annotations can lead physicians astray.[9–12] These findings suggest that automated computer-generated interpretation influences final ECG interpretation. Unfortunately, routinely implemented interpretation algorithms are notoriously flawed and thus pose the risk of patient harm.[4–8] Hence, improving automated interpretation accuracy is an important step toward delivering safe patient care.

Recent studies demonstrate the potential value of deep neural networks for ECG analysis. However, these findings have been limited to single-lead ECGs[12] or fall short in providing complete 12-lead ECG interpretation.[13,14] Our AI-ECG algorithm is capable of comprehensive 12-lead ECG interpretation consistent with those provided by board-certified cardiologists on nearly 2.5 million standard 12-lead ECGs from over 720,000 adult patients. In recent work, we demonstrated that an AI-ECG algorithm could generate 66 structured diagnostic codes from a spectrum of uncommon and complex to normal ECG features (eg, primary and secondary rhythms, axis deviation, chamber enlargement/hypertrophy, atrioventricular and intraventricular conduction delay, myocardial ischemia, waveform abnormalities, clinical disorders, and pacemaker activity).[16]

In this work, we developed a novel AI-ECG algorithm capable of comprehensive but parsimonious, human-like 12-lead ECG interpretation and demonstrated its promising diagnostic performance against currently implemented ECG interpretation methods.

This comparison trial demonstrates that AI-ECG algorithms hold tremendous clinical potential in replacing conventional automated models. Similar to computer-generated interpretations, AI-ECG predictions provide unbiased, reproducible results. However, unlike conventional computer-generated programs, an AI-ECG algorithm can continue to learn and improve its recognition of various patterns in an automatic manner by being fed expert, human-revised interpretations. Therefore, ongoing training of an AI-ECG algorithm on high-quality raw ECG signals could further refine and enhance prediction accuracy to the level that is nearly always considered clinically acceptable. Additional training of existing and development of new algorithms on various regions, populations, and diseases may help improve its prediction accuracy and ensure generalizability.

In current clinical practice, ECG interpretation workflow is highly resource intensive and difficult to scale. An AI-ECG algorithm capable of 12-lead ECG interpretation could help scale the scope of our practice as well as improve consistency and overall accuracy of results. Given the known limitations and notorious inconsistent performance of conventional automated algorithms, AI-ECG has the potential to cause a paradigm shift in preliminary ECG interpretation methods. Although AI-ECG algorithms are unlikely to replace final expert annotation, an AI-ECG algorithm capable of accurate and consistent 12-lead ECG interpretation predictions could optimize clinical workflow by triaging and providing warning of urgent cases, as well as prioritizing and structuring ECG review by clinicians. Ideally, an AI-ECG algorithm would serve as an adjunct to improve interpretation accuracy. Additionally, advances in technology may provide a role for its use in telemedicine, especially in regions where experienced interpreters and resources are limited.

Limitations

The AI-ECG algorithm was derived on randomly selected patients and ECGs from the Mayo Clinic ECG laboratory database and thus, the representativeness of this sample may vary in comparison to other populations. As such, the AI-ECG algorithm may not reflect all ethnic and racial groups, thereby affecting predictions. Further study is needed to evaluate the AI-ECG algorithm's performance in real time and in diverse, population-specific datasets. While no ECGs were included in the initial training and testing datasets of the AI-ECG algorithm, some patients contributed ECGs to each group, which may have made some features more easily recognized and potentially falsely improved the algorithm's performance. This does not eliminate the potential for bias, although the use of 720,978 patients in the initial work likely mitigated this risk. An additional limitation of the AI-ECG algorithm

is the use of error-prone over-reading cardiologist interpretation as the "gold standard" for algorithm development.

Multiple studies have demonstrated that erroneous codes generated by the computer may be propagated and end up as the final clinical interpretation.[9–11] Whether this is owing to the over-reading providers' large work burden and fatigue, it may be the cause for the average of 6.0% of final clinical interpretations reported as inaccurate (score of 3). We did not assess who were the original over-reading cardiologists of all the ECGs used in the study. We believe that there would be at least 20 different over-reading practicing cardiologists that made final interpretations during the time frame of the ECGs selected, and given the large volume of ECGs included, we do not believe that this or their experience would have a significant impact on the study results.

Our study was limited by the number of cardiologists who volunteered to participate. However, we felt that 3 expert cardiac electrophysiology interpreters as well as the inclusion of a large number of ECGs with a wide variety of ECG interpretations would suffice. We believe that future comparison trials should strive for a larger number of ECGs and expert interpreters to examine if this would alter results.

A final and notable limitation of our study was the use of a single computerized ECG interpretation system. We chose this system because it is the one implemented at our institution and the one used to provide the initial interpretation for the over-reading cardiologists upon which the AI-ECG algorithm was developed. It also makes up the majority of the market in ECG acquisition, analysis, and storage. Future study assessing the accuracy of the AI-ECG algorithm against multiple conventional algorithms would help to validate our findings. However, comparison may prove difficult given the lack of consistency of labels across all computer algorithms.

A next step would be to test the AI-ECG algorithm in a controlled true clinical scenario to demonstrate that accurate predictions could be made in real time for subsequent final cardiologist over-read (ie, essentially acting as a substitute to the computer model). This prospective implementation and analysis would allow for instant expert feedback that could be incorporated to improve the algorithm's predictions. Additionally, testing our AI-ECG algorithm on established ECG databases could help assess and affirm its performance in various populations and against other developed models. If an AI-ECG algorithm could be refined so well to perform at the level of a trained cardiologist, it may allow for more complex or urgent cases to require expert review. This would provide tremendous value to low- and middle-income regions where resources are scarce.

Conclusion

We demonstrate that an AI-ECG algorithm outperforms a clinically implemented computer program and better approximates expert cardiologist over-read for comprehensive standard 12-lead ECG interpretation. These results suggest an AI-ECG algorithm may serve as a nonbiased means to

improve interpretation accuracy, optimize workflow, and expand access in resource-limited regions. Further study is warranted to assess the AI-ECG algorithm's accuracy across different populations as well as its application in real-time clinical practice.

Funding Sources

This project was conceived, executed, and funded by Mayo Clinic without industry support.

Disclosures

PAN, ZA, PAF, and Mayo Clinic are involved in potential equity/royalty relationship with AliveCor. ZA, PAF, and Mayo Clinic are involved in potential equity/royalty relationship with Eko. PAN, ZA, PAF, and Mayo Clinic have filed/planned patents related to the application of AI to the ECG for diagnosis of various cardiac conditions.

Ethics Statement

The authors designed the study and gathered and analyzed the data according to the Helsinki Declaration guidelines on human research. The research protocol used in this study was reviewed and approved by the Mayo Clinic institutional review board.

Patient Consent

All patients involved in this research have provided informed consent to partake.

Authorship

All author attest they meet the current ICMJE criteria for authorship.

Disclaimer

Given his role as Section Editor, Zachi Attia had no involvement in the peer review of this article and has no access to information regarding its peer review.

References

1. Taback L, Marden E, Mason HL, Pipberger HV. Digital recording of electrocardiographic data for analysis by a digital computer. IRE Trans Med Electron 1959; 6:167–171.
2. Rikli AE, Tolles WE, Steinberg CA, et al. Computer analysis of electrocardiographic measurements. Circulation 1961;24:643–649.
3. Caceres CA, Steinberg CA, Abraham S, et al. Computer extraction of electrocardiographic parameters. Circulation 1962;25:356–362.
4. Schlapfer J, Wellens HJ. Computer-interpreted electrocardiograms: benefits and limitations. J Am Coll Cardiol 2017;70:1183–1192.
5. Shah AP, Rubin SA. Errors in the computerized electrocardiogram interpretation of cardiac rhythm. J Electrocardiol 2007;40:385–390.
6. Guglin ME, Thatai D. Common errors in computer electrocardiogram interpretation. Int J Cardiol 2006;106:232–237.
7. Poon K, Okin PM, Kligfield P. Diagnostic performance of a computer-based. ECG rhythm algorithm. J Electrocardiol 2005;38:235–238.
8. Bogun F, Anh D, Kalahasty G. Misdiagnosis of atrial fibrillation and its clinical consequences. Am J Med 2004;117:636–642.
9. Martínez-Losas P, Higueras J, Gómez-Polo JC, et al. The influence of computerized interpretation of an electrocardiogram reading. Am J Emerg Med 2016; 34:2031–2032.
10. Lindow T, Kron J, Thulesius H, et al. Erroneous computer-based interpretations of atrial fibrillation and atrial flutter in a Swedish primary health care setting. Scand J Prim Health Care 2019;37:426–433.
11. Anh D, Krishnan S, Bogun F. Accuracy of electrocardiogram interpretation by cardiologists in the setting of incorrect computer analysis. J Electrocardiol 2006;39:343–345.
12. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019;25:65–69.
13. Novotny T, Bond R, Androsva I, 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 2017;101:1183–1192.
14. Ribeiro AH, Ribeiro MH, Paixao GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nat Commun 2020;11:1760.
15. Zhu H, Cheng C, Yin H, et al. Automatic multilabel electrocardiogram diagnosis of heart rhythm or conduction abnormalities with deep learning: a cohort study. Lancet Digit Health 2020;2:e348–e357.
16. Kashou AH, Ko W-Y, Attia ZI, Cohen MS, Friedman PA, Noseworthy PA. A comprehensive artificial intelligence-enabled electrocardiogram interpretation program. Cardiovasc Digit Health J 2020;2:1–9.
17. Salvador A, Drozdzal M, Giro-i-Nieto X, Romero A. Inverse cooking: recipe generation from food images. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2019;10453–10462.
18. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. arXiv:1706.03762. 2017.
19. Paszke A, Gross S, Massa F, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. 2019. https://arxiv.org/abs/1912.01703.
20. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. https://www.R-project.org/.
21. Gamer M, Lenon J, Fellows I, Singh P. Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1. https://CRAN.R-project.org/package=irr. 2019.