

ORIGINAL ARTICLE

Artificial intelligence to detect abnormal heart rhythm from scanned electrocardiogram tracings

Joshua Bridge¹  | Lu Fu² | Weidong Lin² | Yumei Xue² | Gregory Y. H. Lip³  | Yalin Zheng^{1,3} 

¹Department of Eye and Vision Science, Institute of Life Course and Medical Sciences, University of Liverpool, Liverpool, UK

²Department of Cardiology, Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China

³Liverpool Centre for Cardiovascular Science, University of Liverpool and Liverpool Heart & Chest Hospital, Liverpool, UK

Correspondence

Yalin Zheng, William Henry Duncan Building, 6 West Derby Street, Liverpool L7 8TX, UK.
Email: yalin.zheng@liverpool.ac.uk

Funding information

Engineering and Physical Sciences Research Council Studentship (No. 2110275, EP/R513271/1), UK. The National Key Research and Development Program of China (No. 2018YFC1312500, No. 2018YFC1312502), China. The National Natural Science Foundation (No. 81870254), China.

Abstract

Background: Electrocardiogram (ECG) interpretation is an integral part of the clinical ECG workflow; however, this process is often time-consuming and labor-intensive. We aim to develop a rapid, inexpensive means to detect abnormal ECGs using artificial intelligence (AI) from scanned ECG printouts.

Methods: The study included 1172 12-lead ECG scans performed in 1172 individuals from a community in Guangzhou, China; 878 (74.9%) were diagnosed with sinus rhythm, and the remaining 294 (25.1%) with abnormal rhythms. A deep learning model consisting of a convolutional neural network based on InceptionV3 and a fully connected layer followed by a GEV activation was trained to classify scanned tracings as either normal or abnormal.

Results: In a hold-out testing set, the model achieved a area under curve (AUC), sensitivity, specificity, PPV, and NPV of 0.932 (95% confidence interval [CI]: 0.890, 0.976), 0.816 (95% CI: 0.657, 0.923), 0.993 (95% CI: 0.959, 1.0), 0.969 (95% CI: 0.838, 0.999), and 0.950 (95% CI: 0.90, 0.980) respectively, when using a probability threshold of 0.5. When compared with a physiological expert, these results show comparable performance with a statistically significant increase in specificity and a non-significant decrease in sensitivity at the 95% level.

Conclusions: We have developed a rapid, inexpensive, accurate means to detect abnormal ECGs using AI. Easy and accurate identification of such "abnormal" ECGs could allow the mass automated review of ECGs in community settings where abnormal ones could be flagged using AI for detailed clinical review by healthcare professionals.

KEYWORDS

deep learning, ECG, screening

1 | INTRODUCTION

Interpretation of the computerized electrocardiogram (ECG) is an integral component of the clinical ECG workflow.¹ As a result of a large

amount of subjectivity in the interpretation and considerable noise within the signal, the ECG may be subtly altered when diagnosed.² Inexperienced physicians ordering the ECG may fail to recognize interpretation mistakes made by the automated diagnosis algorithm

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Journal of Arrhythmia* published by John Wiley & Sons Australia, Ltd on behalf of the Japanese Heart Rhythm Society.

within the machine used for recordings and accept the automatic diagnosis without criticism. A wrong diagnosis increases the risk of exposing patients to unnecessary investigations or potentially dangerous treatment. Comparisons of computer-based and expert reader performance are likely to show evolving results with continuous improvement of computer-ECG algorithms and ECG interpreters' changing expertise.¹

Artificial intelligence (AI), in particular deep learning (DL), has previously been utilized to alleviate the pressure on clinicians by automatically diagnosing a variety of diseases in a variety of fields, including cardiology,³ ophthalmology,⁴ and neurology.⁵ Given the high cost of ECG community screening⁶ and the shortage of trained professionals needed to read and diagnose obtained ECG recordings, automated algorithms have been used.

Computerized interpretation of ECGs (CIE) has provided the automatic diagnosis of ECG signals saving considerable time for clinicians; however, this relies on ECG machines being equipped with this technology, which is often proprietary and differs between devices. Another limitation of CIE is that the results are often "black-box" with little to no indication to the clinician as to why the diagnosis was made.

With an aging global population, applying an efficient and accurate automated ECG interpretation algorithm can help provide convenient solutions for diagnosis in community screening settings; this could improve the early detection of abnormal ECGs.

We hypothesized that DL can accurately classify scanned ECG graphs as either normal or abnormal. To this end, we developed an ECG system-agnostic DL model to detect abnormal scanned ECG graphs automatically. We aim to classify scanned ECG printouts as "normal" or 'abnormal' in a community screening setting, where patients display no known symptoms of cardiovascular disease. Our method provides a method to visualize why the classification was made through saliency maps,⁷ highlighting the areas of the ECG graph used in the final classification.

Hence, we aimed to develop and validate a quantitative method to detect abnormal ECG graphs from scanned graphs using AI with "real" 12-lead ECG images directly scanned from printouts. This is important in clinical settings since the most common output is a 12-lead ECG printout.

2 | MATERIALS AND METHODS

2.1 | Dataset

The Guangzhou Heart Study⁸ provided the data underlying this article by permission and cannot be shared publicly. We included digital 1172 ECGs performed for 1172 individuals screened from May 1, 2019, to August 31, 2019, in Guangzhou, China, which was the second phase of the Guangzhou Heart Study.⁸ Included patients were permanent residents of Guangzhou aged 35 or over living in the sample community for at least 6 months. Patients were excluded if they¹ had mental or cognitive disorders,² mobility issues,³ were

pregnant or lactating,⁴ were under treatment for malignant tumors,⁵ were temporary residents,⁶ had not lived in the sample community for at least 6 months,⁷ were non-Guangzhou residents, or⁸ did not respond. The paper ECGs obtained at the site were scanned into digital files for storage. The diagnosis of the ECGs was made by two specific electrophysiological experts.⁸ Abnormal ECGs included: sinus arrhythmia, atrial fibrillation (AF) and flutter (AFL), premature atrial or ventricular contraction, atrioventricular block (AVB), ventricular tachycardia, supraventricular tachycardia (SVT), Wolff-Parkinson-White syndrome (WPW), pacing rhythm, and junctional rhythm. Here, sinus arrhythmia is considered abnormal, as the main aim is to develop an algorithm for community screening, which is considered useful to detect sinus arrhythmia.

The paper records were scanned at 300 dpi in both horizontal and vertical directions; the generated RGB images were cropped to 2900x1700 pixels to exclude patients' details such as name and date of birth which appear on the top of the ECG printout. None of the obtained images were excluded from the study.

All 1172 ECGs performed for 1172 individuals (mean age 71.4 years [standard deviation 6.3]; 34% male) were eligible for this study. Of 1172 ECGs, 878 (74.9%) were diagnosed with sinus rhythm, and the remaining 294 (25.1%) were deemed "abnormal" by expert consensus.

We split data into training (800 images), validation (200 images), and testing (172 images) sets. The algorithm was trained iteratively on the training set, and the best model was chosen using the validation dataset. Finally, we evaluated the selected model on the testing dataset. Patient characteristics of each data split are presented in [Table 1](#).

2.2 | Preprocessing

Images were first downsampled to 512 x 256 pixels using the Lanczos sampling algorithm to reduce computational complexity while maintaining a similar aspect ratio to the original image. We

TABLE 1 Data. Patient characteristics of each of the data splits, training validation, and testing

	Training/ validation	Testing
Number of patients (% abnormal)	1000 (20%)	172 (22%)
Sex		
Female	645 (64.5%)	128 (74.4%)
Male	355 (35.5%)	44 (25.6%)
Age (years)		
Mean	71.4	71.5
Standard deviation	6.3	6.0
Median	70	70
Min	49	60
Max	96	91

then normalized the images to between 0 and 1. During training, we rotated each image by -10 to 10 degrees. We also randomly altered the brightness to between 80 and 120%, allowing the algorithm to better generalize to new unseen images.

2.3 | Classification

We classify the images as normal or abnormal using a convolutional neural network (CNN). The network is based on Inception V3,⁹ a popular and highly generalizable network developed by Google. Inception V3 uses regularization and factorized kernels, resulting in an accurate yet computationally efficient network. The network was pretrained on Imagenet,¹⁰ a collection of natural images such as dogs, cats, and plants. Pretraining reduces the time needed for training as the network has already learned some semantic features in other images. As a result, the network only needs to be fine-tuned on new images rather than learning those features from scratch; this concept is known as transfer learning. Inception V3⁹ results in a single vector representation of each image of length 2048 using a pooling layer; this network is outlined in Figure 1. The vector representation was then classified using a dense, fully connected layer. After the pooling layer, we added a dropout layer and an L1 and L2 regularization of $1e-5$ on the fully-connected layer to reduce overfitting and improve generalizability without increasing computational complexity.

As a result of the unbalanced class distribution, we chose the generalized extreme value (GEV) activation function as the final activation.¹¹ The GEV activation is an alternative to the commonly

used sigmoid activation and has been shown to improve performance with unbalanced data. This results in a predicted probability of the scanned ECG printout displaying an abnormal rhythm.

2.4 | Computing

All analyses were conducted on a Linux machine running Ubuntu 18.04, with a Titan X 12GB GPU and 32GB of memory. We developed the DL model using Tensorflow. We used a weighted binary cross-entropy for the loss function and the Adam optimizer with an initial learning rate of 0.001. If the validation loss had not improved within three epochs, the learning rate was reduced to two-thirds. Model checkpoints and early stopping with a patience of 7 were used to prevent overfitting. The best model was selected based on the weighted binary cross-entropy in the validation set. Code is available at <https://github.com/JTBridge/ECG>.

2.5 | Metrics

We assessed model performance using the Brier score,¹² the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Confidence intervals were built using DeLong's method¹³ for the AUC and exact binomial confidence limits for the sensitivity, specificity, PPV, and NPV. All metrics were calculated in R using the pROC¹⁴ and epiR¹⁵ packages.

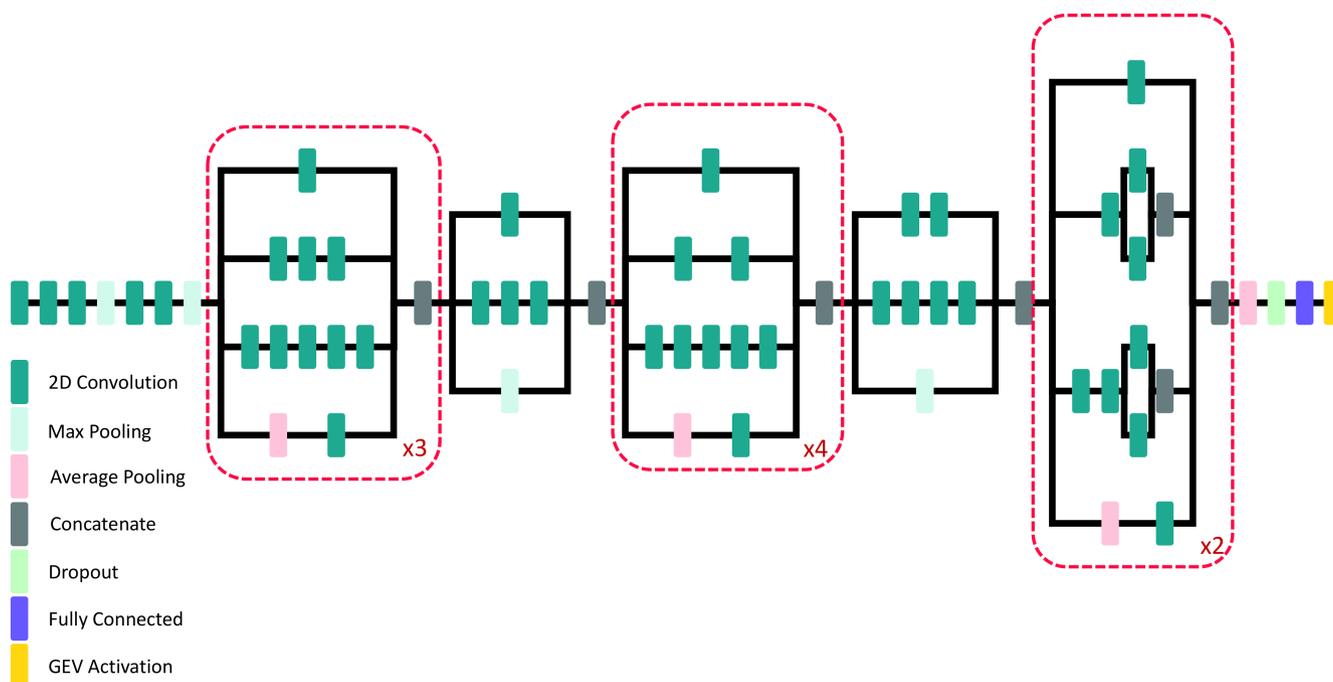


FIGURE 1 Inception V3. Diagram of InceptionV3 network architecture, showing multiple layers

2.6 | Saliency maps

To visualize how the network reached each decision, we produced saliency maps,⁷ highlighting the areas that the algorithm believes to be important in the classification. These maps can be used to justify the classification to health professionals.

Saliency maps were produced using guided backpropagation. Backpropagation shows which parts of the image are most important in deciding the final classification by assessing which pixels affect the neurons the most. The gradient of each neuron is calculated, and the neurons with the highest gradients are the most activated. Guided backpropagation extends this by setting any negative gradients to zero. The most activated neurons can then be formed into a heat map and overlaid to determine which areas of the image are most important.

3 | RESULTS

We fine-tuned the Inception V3 network, pretrained on ImageNet, on 800 ECG images. The best model was chosen based on the

validation binary cross-entropy loss on 200 images. The model was then tested on another 172 images in the hold-out testing dataset.

On the testing dataset, the final model achieved an area under the receiver operating characteristic (AUC) of 0.932 (95% CI: 0.890, 0.976) and sensitivity, specificity, PPV, and NPV of 0.816 (95% CI: 0.657, 0.923), 0.993 (95% CI: 0.959, 1.0), 0.969 (95% CI: 0.838, 0.999), and 0.950 (95% CI: 0.90, 0.980) respectively, with a probability threshold of 0.5. Full results for a range of probability thresholds are shown in Table 2, with the ROC curve shown in Figure 2.

We compared our results with those of a cardiologist who specializes in electrophysiology. The 12-lead ECG interpretation through the final model showed high consistency with the electrophysiologist's diagnosis. The sensitivity, specificity, PPV, and NPV of the expert's predictions on the testing dataset were 0.947 (95% CI: 0.842, 0.988), and 0.918 (95% CI: 0.862, 0.955), 0.766 (95% CI: 0.632, 0.869), 0.984 (95% CI: 0.950, 0.997), respectively. At a probability threshold of 0.5, our model shows a statistically significant increase in specificity and PPV with a non-significant decrease in specificity and NPV at the 95% confidence level. We obtained a prevalence and

TABLE 2 Results on the testing dataset for the human expert and the deep learning model

Method	Brier Score	AUC	Probability threshold	Sensitivity	Specificity	PPV	NPV
Ours	0.0705	0.932 (0.890,0.976)	0.3	0.868 (0.719, 0.956)	0.993 (0.959, 1.0)	0.971 (0.847, 0.999)	0.964 (0.917, 0.988)
			0.4	0.842 (0.687, 0.940)	0.993 (0.959, 1.0)	0.970 (0.842, 0.999)	0.957 (0.908, 0.984)
			0.5	0.816 (0.657, 0.923)	0.993 (0.959, 1.0)	0.969 (0.838, 0.999)	0.950 (0.90, 0.980)
			0.6	0.816 (0.657, 0.923)	0.993 (0.959, 1.0)	0.969 (0.838, 0.999)	0.950 (0.90, 0.980)
			0.7	0.789 (0.627, 0.904)	1.0 (0.973, 1.0)	1.0 (0.884, 1.0)	0.944 (0.892, 0.975)
Expert	0.0363	—	—	0.947 (0.823, 0.990)	0.918 (0.858, 0.958)	0.766 (0.620, 0.877)	0.984 (0.943, 0.998)

DeLong's method is used to calculate the 95% confidence intervals for the AUC, and exact binomial confidence intervals are used for sensitivity, specificity, positive predictive value, and negative predictive value.

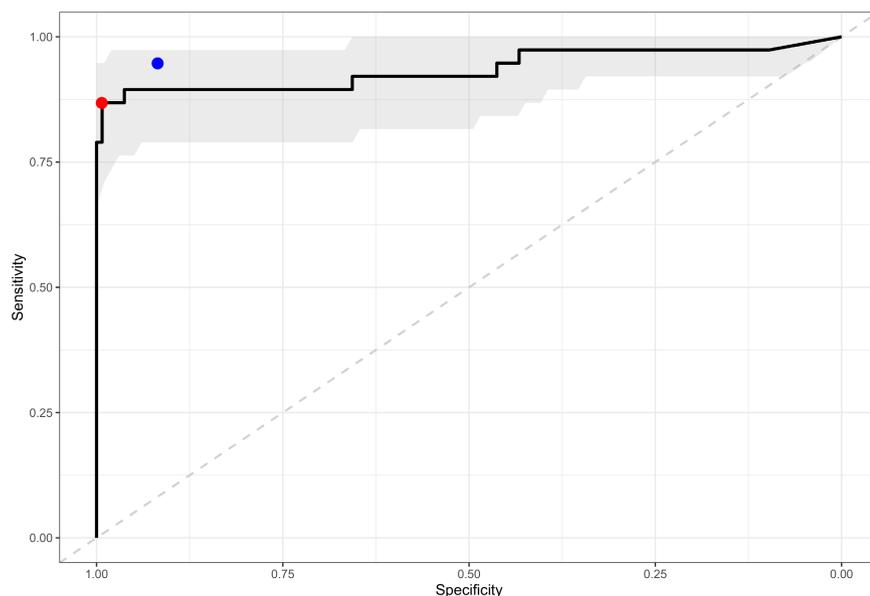


FIGURE 2 ROC Curve. The receiver operating characteristic curve shows the testing dataset's discrimination performance, with a 95% confidence band shown in grey. The area under the curve is 0.935 (0.871, 0.999). The sensitivity and specificity are shown as a red point. The expert's performance is shown with a blue point

bias-corrected kappa score of 0.757, indicating substantial agreement between the expert and the algorithm. Results with comparisons are summarised in Table 2.

Examples of ECGs and their saliency maps are shown in Figure 3. The saliency map for the normal ECG shows little activation; this image was predicted with a 0% probability of being abnormal. The abnormal ECG saliency map highlights abnormal parts; the algorithm predicted that this was abnormal with a probability of 99.6%. The final saliency map shows an image incorrectly classified as abnormal, although the algorithm was not entirely sure, with a probability of only 64.1%. The incorrect classification appears to be caused by some interference in the signal.

4 | DISCUSSION

In this study, we have demonstrated the ability of DL to detect abnormal ECGs automatically from scanned printouts. By fine-tuning a pretrained DL network, we can accurately classify scanned ECGs as normal or abnormal. Easy and accurate identification of such "abnormal" ECGs could allow the mass automated review of ECGs where abnormal ones could be flagged up using AI for detailed clinical review by healthcare professionals.

We obtained a final testing AUC of 0.932 (95% CI: 0.890, 0.976), and an optimal sensitivity and specificity of 0.816 (95% CI: 0.657, 0.923) and 0.993 (95% CI: 0.959, 1.0), respectively. Results show no

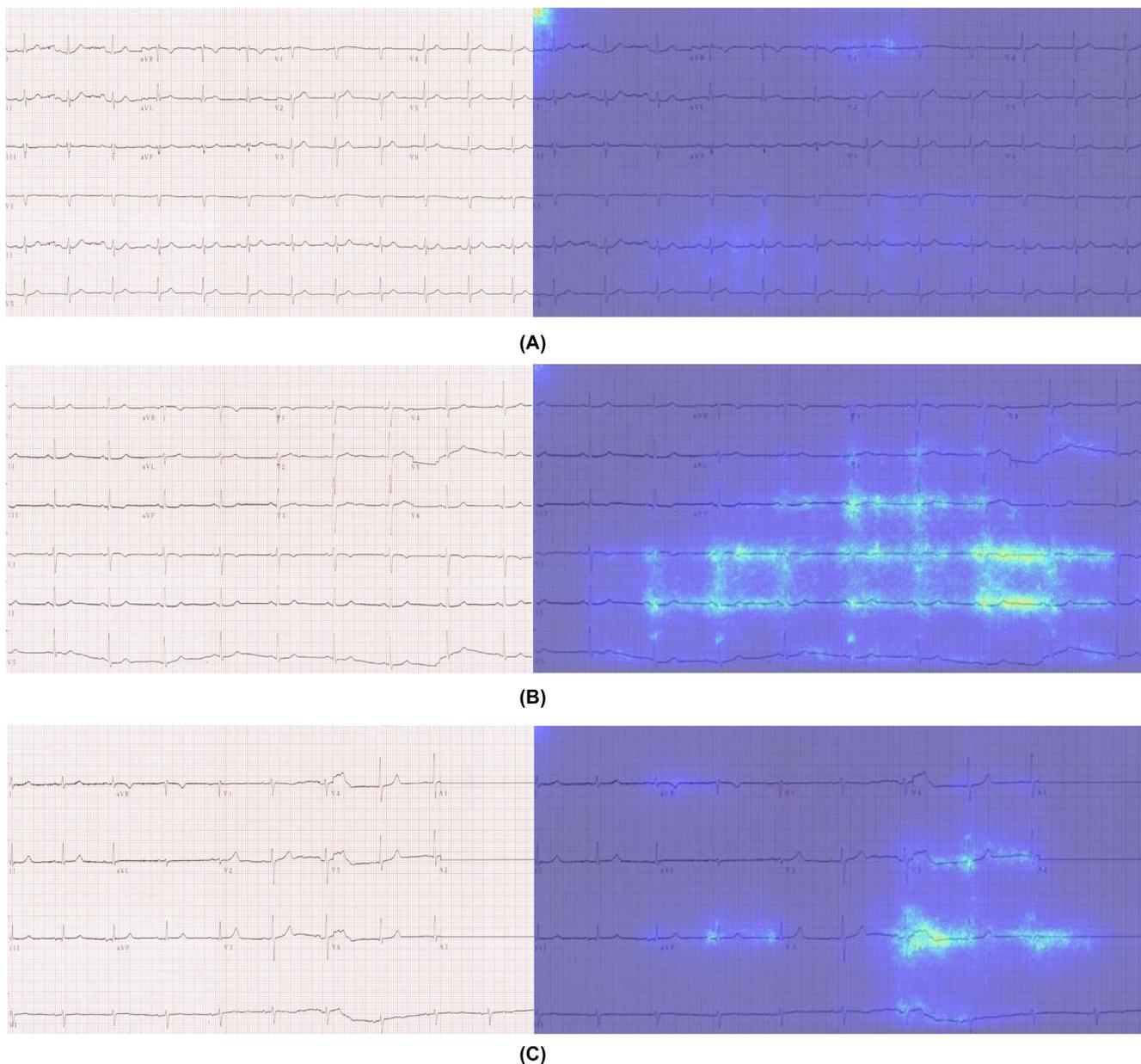


FIGURE 3 Saliency maps. Examples of scanned ECG images and their corresponding saliency maps showing: (A) correctly identified normal ECG, (B) correctly identified abnormal ECG, (C) normal ECG wrongly identified as abnormal. Brighter areas show areas of the image that the algorithm finds most useful in the classification

statistically significant difference compared to an electrophysiologist, except for specificity, which shows a significant increase.

Indeed, these results show that the algorithm can correctly identify abnormal ECG scans with performance similar to human experts. This could potentially be utilized in a community screening setting, where a trained clinician may not be available to interpret ECGs. The algorithm may also be used as a secondary check mechanism to ensure an abnormal ECG is not missed.

The proposed algorithm provides a fast, accurate, system-agnostic method to classify ECG graphs. While previous methods rely on analyzing the raw waveform signal, our approach uses the scanned ECG paper directly without the need to extract the raw waveform. This approach is advantageous in community and low-resource settings. In addition, the algorithm's saliency map helps to justify the classification to clinicians so that the diagnosis may be checked for mistakes. In clinical settings, the most common output is a 12-lead ECG printout. Nonetheless, we are aware that ECG records can also be exported in PDF files and would be much easier to process as the signal-noise ratio is much higher than the images digitized from the paper record.

A previous study by Attia et al.¹⁶ reported accurate classification of raw single-lead ECG signals using a DL algorithm; this approach requires access to the raw signals and cannot be used on ECG graph printouts. They describe an AI-enabled ECG algorithm for the identification of patients with AF during sinus rhythm, which showed that a single AI-enabled ECG identified AF with an AUC of 0.87 (0.86–0.88), a sensitivity of 0.79 (77.5–80.4), a specificity of 0.795 (0.790–0.799), and overall accuracy of 0.794 (0.790–0.799). Hannun et al.¹⁷ showed that DL can provide comparable performance to cardiologists in classifying arrhythmias from single-lead ECGs. This could be used to prioritize the most severe patients in a triage setting. Sengupta et al.¹⁸ used machine learning techniques to diagnose early diastolic dysfunction from a 12-lead ECG. They collected a sample of 188 patients' ECGs who were referred for coronary computed tomography and found that the machine learning techniques demonstrated good sensitivity (0.80) and specificity (0.84) for diagnosing early diastolic dysfunction, with an AUC of 0.91 (0.86–0.95) for prediction of abnormal myocardial mechanical relaxation. Using only 2 ECG leads from a 12-lead ECG, Galloway et al.¹⁹ developed a DL model to detect hyperkalemia in patients with the renal disease with AUCs of 0.853 to 0.883.

How would AI compare to healthcare professional reporting? Ribeiro et al.²⁰ concluded that the Deep Neural Networks outperformed cardiology resident medical doctors in recognizing six types of abnormalities in 12-lead ECG recordings, with F1 scores above 80% and specificity over 99%.

In clinical settings, ECGs are commonly provided in the form of a printout. For digital signal processing such as AI, this printout must be digitized. There have been some efforts in extracting raw ECG signals after digitizing the paper ECG record but currently published studies have a small sample size. For example, Baydoun et al.²¹ showed excellent performance on 30 ECG scanned available curves. However, Patil et al.²² only used nine paper records.

Why the limited data on using digitized ECGs for AI studies? Waits et al.²³ highlighted challenges facing digitizing ECG, noting that there

are no techniques that can readily be used for digitization tasks. A large amount of historical data will not be analyzed by previous DL methods requiring ECG waveform signals. Indeed, Brisk et al.²⁴ demonstrated the feasibility of using DL to automatically interpret ECG images, producing "synthetic" ECG printouts from ECG wave signals, and applying DL for the classification. However, their generated images were free of noise, and only signals from a single lead were used.

5 | LIMITATIONS

This study's main limitation is the relatively small sample size, with only 172 patients in the testing data. This proof of concept method needs to be externally validated, either temporally or spatially, to confirm the utility of DL in this manner. Second, this study only considers sinus rhythm against abnormal rhythm; this may be useful in community screening programs; however, it would be more beneficial to diagnose the type of abnormal rhythm detected. Future work may wish to consider the differential diagnosis to distinguish between the many types of abnormal ECG patterns. Further clinical information such as age and comorbidities may also be beneficial in the diagnosis and could be added to the model.

6 | CONCLUSIONS

Deep learning can be used to detect abnormal ECG graphs with human-level performance, thus reducing pressure on healthcare professionals. Saliency maps can be used to justify the classification when making clinical decisions. Easy and accurate identification of such "abnormal" ECGs could allow the mass automated review of ECGs where abnormal ones could be flagged for detailed clinical review.

ACKNOWLEDGMENTS

We thank Dr Hai Deng and Dr Shulin Wu for their work in collecting data and maintaining the Guangzhou database.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

The dataset analyzed in this study is taken from the Guangzhou Heart Study [<https://doi.org/10.1038/s41598-018-35928-w>].

ORCID

Joshua Bridge  <https://orcid.org/0000-0001-8798-4134>

Gregory Y. H. Lip  <https://orcid.org/0000-0002-7566-1626>

Yalin Zheng  <https://orcid.org/0000-0002-7873-0922>

REFERENCES

1. Rautaharju PM. Eyewitness to history: Landmarks in the development of computerized electrocardiography. *J Electrocardiol*. 2016;49(1):1–6.

2. Jahmunah V, Oh SL, Wei JKE, Ciaccio EJ, Chua K, San TR, et al. Computer-aided diagnosis of congestive heart failure using ECG signals - A review. *Phys Med*. 2019;62:95–104.
3. Miyazawa AA. Artificial intelligence: the future for cardiology. *Heart*. 2019;105(15):1214.
4. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2):167–75.
5. Patel UK, Anwar A, Saleem S, Malik P, Rasul B, Patel K, et al. Artificial intelligence as an emerging technology in the current care of neurological disorders. *J Neurol*. 2021;268:1623–42.
6. Halkin A, Steinvil A, Rosso R, Adler A, Rozovski U, Viskin S. Preventing sudden death of athletes with electrocardiographic screening: what is the absolute benefit and how much will it cost? *J Am Coll Cardiol*. 2012;60(22):2271–6.
7. Smilkov D, Thorat N, Kim B, Viégas F, Wattenberg M. Smoothgrad: removing noise by adding noise. arXiv preprint arXiv:170603825. 2017.
8. Deng H, Guo P, Zheng M, Huang J, Xue Y, Zhan X, et al. Epidemiological characteristics of atrial fibrillation in Southern China: results from the Guangzhou Heart Study. *Sci Rep*. 2018;8(1):17829.
9. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z, editors. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
10. Deng J, Dong W, Socher R, Li L, Kai L, Li F-F, editors. ImageNet: A large-scale hierarchical image database. IEEE Conference on Computer Vision and Pattern Recognition; 2009.
11. Bridge J, Meng Y, Zhao Y, Du Y, Zhao M, Sun R, et al. Introducing the GEV activation function for highly unbalanced data to develop COVID-19 diagnostic models. *IEEE J Biomed Health Inform*. 2020;24(10):2776–86.
12. Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev*. 1950;78(1):1–3.
13. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837–45.
14. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform*. 2011;12(1):77.
15. Stevenson M, Sergeant E, Nunes T, Heuer C, Marshall J, Sanchez J, et al. epiR: Tools for the Analysis of Epidemiological Data. 2022.
16. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet*. 2019;394(10201):861–7.
17. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. 2019;25(1):65–9.
18. Sengupta PP, Kulkarni H, Narula J. Prediction of abnormal myocardial relaxation from signal processed surface ECG. *J Am Coll Cardiol*. 2018;71(15):1650–60.
19. Galloway CD, Valys AV, Shreibati JB, Treiman DL, Petterson FL, Gundotra VP, et al. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. *JAMA Cardiol*. 2019;4(5):428–36.
20. Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun*. 2020;11(1):1760.
21. Baydoun M, Safatly L, Abou Hassan OK, Ghaziri H, El Hajj A, Isma'eel H. High precision digitization of paper-based ecg records: a step toward machine learning IEEE J Transl Eng Health Med 2019;7:1900808.
22. Patil R, Karandikar R. Digitization and Parameter Extraction of Preserved Paper Electrocardiogram Records. *Soft Computing and Signal Processing*. Springer; 2019 487–95.
23. Waits GS, Soliman EZ. Digitizing paper electrocardiograms: Status and challenges. *J Electrocardiol*. 2017;50(1):123–30.
24. Brisk R, Bond R, Banks E, Peadro A, Finlay D, McLaughlin J, et al. Deep learning to automatically interpret images of the electrocardiogram: do we need the raw samples? *J Electrocardiol*. 2019;57:S65–S9.

How to cite this article: Bridge J, Fu L, Lin W, Xue Y, Lip GY, Zheng Y. Artificial intelligence to detect abnormal heart rhythm from scanned electrocardiogram tracings. *J Arrhythmia*. 2022;38:425–431. doi:[10.1002/joa3.12707](https://doi.org/10.1002/joa3.12707)