Artificial intelligence–enabled mobile electrocardiograms for event prediction in paroxysmal atrial fibrillation

Ananditha Raghunath, MS,* Dan D. Nguyen, MD,[†] Matthew Schram, PhD,[‡]
David Albert, MD,[‡] Shyamnath Gollakota, PhD,* Linda Shapiro, PhD,*
Arun R. Sridhar, MBBS, MPH, FHRS[§]

*From the *Department of Computer Science & Engineering, University of Washington, Seattle, Washington,
[†]St. Luke's Mid America Heart Institute, Kansas City, Missouri, [‡]AliveCor, Inc., Mountainview,
California, and [§]University of Washington Heart Institute, Department of Medicine, University of
Washington, Seattle, Washington.*

BACKGROUND Paroxysmal atrial fibrillation (AF) often eludes early diagnosis, resulting in significant morbidity and mortality. Artificial intelligence (AI) has been used to predict AF from sinus rhythm electrocardiograms (ECGs), but AF prediction using sinus rhythm mobile electrocardiograms (mECG) remains unexplored.

OBJECTIVE The purpose of this study was to investigate the utility of AI to predict AF events prospectively and retrospectively using sinus rhythm mECG data.

METHODS We trained a neural network to predict AF events from sinus rhythm mECGs obtained from users of the Alivecor KardiaMobile 6L device. We tested our model on sinus rhythm mECGs within ±0–2 days, ±3–7 days, and ±8–30 days from AF events to determine the optimal screening window. Finally, we tested our model on mECGs from before an AF event to determine whether AF can be predicted prospectively.

RESULTS We included 73,861 users with 267,614 mECGs (mean age 58.14 years; 35% women). Users with paroxysmal AF contributed 60.15% of mECGs. Model performance on the test set comprising control and study samples across all windows of interest showed an area under the curve (AUC) score of 0.760 (95% confidence interval [CI] 0.759–0.760), sensitivity of 0.703 (95% CI 0.700–0.705), specificity of 0.684 (95% CI 0.678–0.685), and accuracy of 69.4% (95% CI 0.692–0.700). Model performance was better on ±0–2 day samples (sensitivity 0.711; 95% CI 0.709–0.713) and worse on the ±8–30 day window (sensitivity 0.688; 95% CI 0.685–0.690), with performance on the ±3–7 day window falling in between (sensitivity 0.708; 95% CI 0.704–0.710).

CONCLUSION Neural networks can predict AF using a widely scalable and cost-effective mobile technology prospectively and retrospectively.

KEYWORDS Artificial intelligence–based electrocardiographic analysis; Atrial fibrillation; Atrial fibrillation event prediction; Mobile electrocardiography; Scalable technology; Sinus rhythm

(Cardiovascular Digital Health Journal 2023;4:21–28) © 2023 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

Atrial fibrillation (AF) is the most common arrhythmia diagnosed in the United States.[1] Early diagnosis of AF is crucial to preventing complications such as stroke and heart failure.[2,3] Diagnosing AF can be challenging because one-third of individuals with the arrhythmia are asymptomatic,[4] and AF often presents intermittently, referred to as *paroxysmal AF*. Strategies for detecting AF include in-hospital monitoring, serial electrocardiography (ECG), and long-term outpatient monitoring using wearable continuous ECGs, event monitors, or implantable cardiac monitors.

Despite these measures, AF detection rates remain low, between 5% and 20%.[5,6] Identifying novel, convenient, and more cost-effective techniques to predict when an AF event may occur could substantially improve early treatment of this condition, which is expected to affect about 12.1 million individuals in the United States by 2030.

Artificial intelligence (AI) algorithms have been harnessed to predict AF development by analyzing health care–based 12-lead ECGs of individuals in sinus rhythm.[7,8] However, 12-lead ECGs might not be available for those who lack an indication for a screening ECG or those without regular access to health care. For these and other cohorts, multiple mobile electrocardiographic (mECG) platforms are emerging that are increasingly being used in physician's offices and by users themselves to monitor heart rhythms.[9–13] The growing ubiquity of mECG devices and the ease of

Address reprint requests and correspondence: Dr Arun R. Sridhar, Division of Cardiology, University of Washington, 1959 NE Pacific St, P.O. Box 356422, Seattle, WA 98195. E-mail address: arun11@cardiology.washington.edu.

KEY FINDINGS

- In this study of more than 260,000 mobile electrocardiogram (mECGs) in 73,861 AliveCor 6L users, we found that neural networks were able to predict AF events from sinus rhythm mECGs with good performance metrics.

- We demonstrated that atrial fibrillation (AF) events can be predicted before they occur. Furthermore, our model shows comparable performance of mECGs obtained before and after the AF event.

- We showed that the accuracy of AF prediction improves when mECGs are screened closer to the AF event, with the predictions on mECGs from the 2-DAY window outperforming the 7-DAY and 30-DAY windows.

- This study is the first to show the utility and feasibility of using artificial techniques on mECGs to predict AF events, which may be a widely scalable technology for reducing morbidity and mortality.

---

self-administration of mECGs with these platforms provide a unique opportunity to leverage AI to perform large scale, cost-effective AF prediction.

In this study, we documented our application of AI models to predict AF events in study participants with paroxysmal AF using self-administered 6-lead sinus rhythm mECGs collected on a smartphone-based mobile device. We studied AF event prediction using sinus rhythm ECGs obtained within different prespecified time windows of the sampling of the AF mECG. Furthermore, we examined whether AF events can be predicted prospectively, which could offer unique therapeutic opportunities.

Methods
Data source and acquisition
Our data consisted of mECGs acquired from KardiaMobile 6-lead (6L) users (AliveCor, Inc., Mountain View, CA) from June 28, 2019, to February 19, 2021. The KardiaMobile 6L device (Figure 1) is an Food and Drug Administration (FDA)–cleared, portable, wireless, 6L mECG platform that can obtain up to 5 minutes of mECG data through 2 touch sensors with no wires, patches, or gels. Each mECG is sampled at 300-Hz resolution, and the mECG devices are configured to record in the bandwidth range of 0.540 Hz (in accordance with, and in excess of, the standard set for ambulatory ECGs described in the IEC 60601-2-47 guidelines).[14]

To record an mECG with KardiaMobile 6L, users place their fingers or thumbs on the top electrodes, place the bottom electrode on the bare skin of their left ankle or left knee, and hold for 30 seconds. If an uninterpretable poor-quality mECG is obtained, the reading is erased, and the user is prompted to resubmit another reading. After mECG acquisi-

tion, data are transmitted through the AliveCor mobile phone application to a cloud-based data repository. Of note, upon registration of a new device, users provided research authorization and self-reported age and sex. Data validity is maintained by trained AliveCor personnel through rigorous auditing. The Institutional Review Board of the University of Washington provided a waiver of informed consent because all study data were de-identified. The research reported in this paper adhered to the Helsinki Declaration as revised in 2013 guidelines.

Model design and implementation
We framed the detection of AF from mECGs as a time series classification problem. The input to our neural network was a 9000 by 2-dimensional matrix, which represented leads I and II sampled at 300 Hz for 30 seconds. Although the KardiaMobile 6L device acquires ECG data in leads I, II, III, aVL, aVR, and aVF, only data in leads I and II were used as input because the network could learn any linear combination of the 4 remaining leads from those 2 leads.

Model architecture consists of 3 sequential parts: a convolutional neural network (CNN); a gated recurrent unit (GRU); and standard dense neural network layers.[15] We used a modified version of the CNN architecture of Attia et al[7] to build our Resnet-based CNN layers. Each Resnet block contained 3 convolutional layers with a fixed filter count and kernel length (Figure 2). Skip connections were used between layers to increase depth and prevent the vanishing gradient. Each Resnet block was followed by a dropout layer to prevent overfitting of the model to the training samples. Our convolutions focused on the longer axis (9000) to extract temporal features, and results were combined across both leads.

We appended a recurrent neural network (RNN) block to the CNN to provide persistence and memory for temporal patterns found across longer patches of signal. We used a bidirectional GRU architecture for our RNN. The GRU, using update and reset gates, also minimized the vanishing gradient problem and decreased the number of parameters, contributing to easy training and reduced overfitting.

The dense layer at the end of the architecture reduced the result from the GRU to a 2-dimensional output vector. A softmax activation function was applied to the output vector, yielding predictive probabilities corresponding to the occurrence or absence of AF. All evaluation and training were done in Tensorflow 2.4.1 (Google, Mountainview, CA) and Python 3.7 on a CentOS-based Linux machine outfitted with a 32 core Intel(R) Xeon(R) CPU E5-2620 @ 2.10 GHz and 4 Nvidia GeForce GTX 1080 GPUs.

Study design
We identified 76,891 users from the pool of all commercial users of the AliveCor KardiaMobile 6L platform who recorded at least one 30-second mECG from the study period. Users were categorized as not having AF (control cohort) if they had no AF readings and had at least 10 sinus rhythm mECG recordings in the Alivecor database.
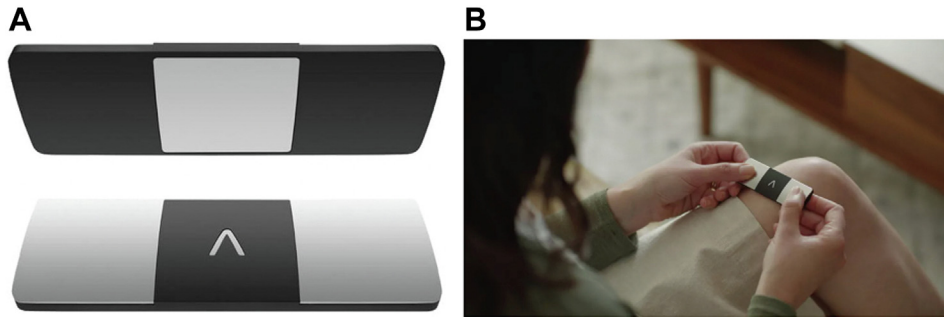
Figure 1   **A:** The Alivecor KardiaMobile 6L mobile electrocardiographic (ECG) device contains 3 electrodes: 2 on the top surface and 1 on the bottom surface. **B:** Six-lead ECG acquisition requires contact with 1 finger from each arm, and from the knee or ankle from a lower limb.

Users were categorized as having AF (study cohort) if they submitted at least 1 mECG with AF, as determined by the proprietary, FDA-cleared Kardia AI algorithm package (FDA clearance K181823).[16] This algorithm uses a mixture of deep learning and conventional machine learning to detect AF from lead I of an ambulatory bandwidth ECG with >95% specificity and 95% sensitivity. Users were excluded if they had persistent or permanent AF (defined, for the study purpose, as AF present in all sample mECGs from an individual) or were <18 years old. mECGs were excluded if they were <30 seconds in duration.

For the primary analysis, we randomly sampled 1 AF mECG from each participant in the study cohort, which served as our cases. Then, from the same participant, we randomly sampled at most 10 sinus rhythm mECGs within ±0–30 days of the sampled AF mECG to include in our study sample. This was done to ensure variety. From each participant in our control cohort, we randomly sampled 1–2 mECGs from the study window to include in our control sample. Users in the study and control cohorts were randomly assigned to either train, validation, or test sets in a 7:1:2 ratio. Therefore, no unique users could have mECGs in >1 of these sets.

In the clinic, we would not have any information about when a patient with suspected paroxysmal AF had the last AF episode. We needed our model to be able to infer AF/no AF without the use of such temporal information. However, we were interested in determining the optimal AF screening window. Therefore, we trained, validated, and tested a single model on study and control mECGs. We reported these overall results, where study samples come from ±0–30 days of an AF event (Figure 3). For identifying an optimal screening window, we further stratified our test set's study mECGs temporally into 3 windows. We chose ±0–2 days (2-DAY), ±3–7 days (7-DAY), and ±8–30 days (30-DAY) windows from the associate AF event because we hypothesized *a priori* that miniscule but clinically important structural and abnormal ECG changes would be most pronounced shortly before, and immediately after, an AF event, and that these changes may be detected by the model. We coupled the control test sample to each of these windows in the reported results, as the control mECGs do not surround a given mECG and rather are sampled from the entire study period.

After exclusions, our final cohort consisted of 73,861 patients and 267,469 mECGs, of which there were 8,661 users who had paroxysmal AF, contributing 160,840 mECGs, and 55,200 control users, contributing 106,629 mECGs. The sampling and stratification methodology is shown in Figure 3. Once control users and study users are assigned to the test set, their mECGs are separated into the different test set subsamples (Figure 4).

We performed secondary analyses to determine the prospective predictive power of our model. We divided the test sample into 2 cohorts: all mECGs collected in the days
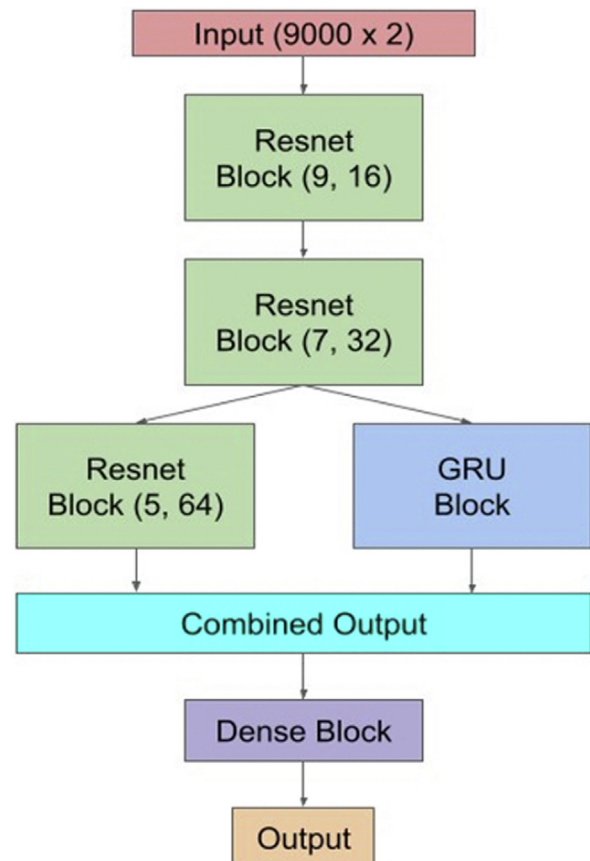

Figure 2   Neural network architecture consisting of Resnet blocks (convolutional neural network [CNN]), gated recurrent unit (GRU) block (RNN), and Dense block.
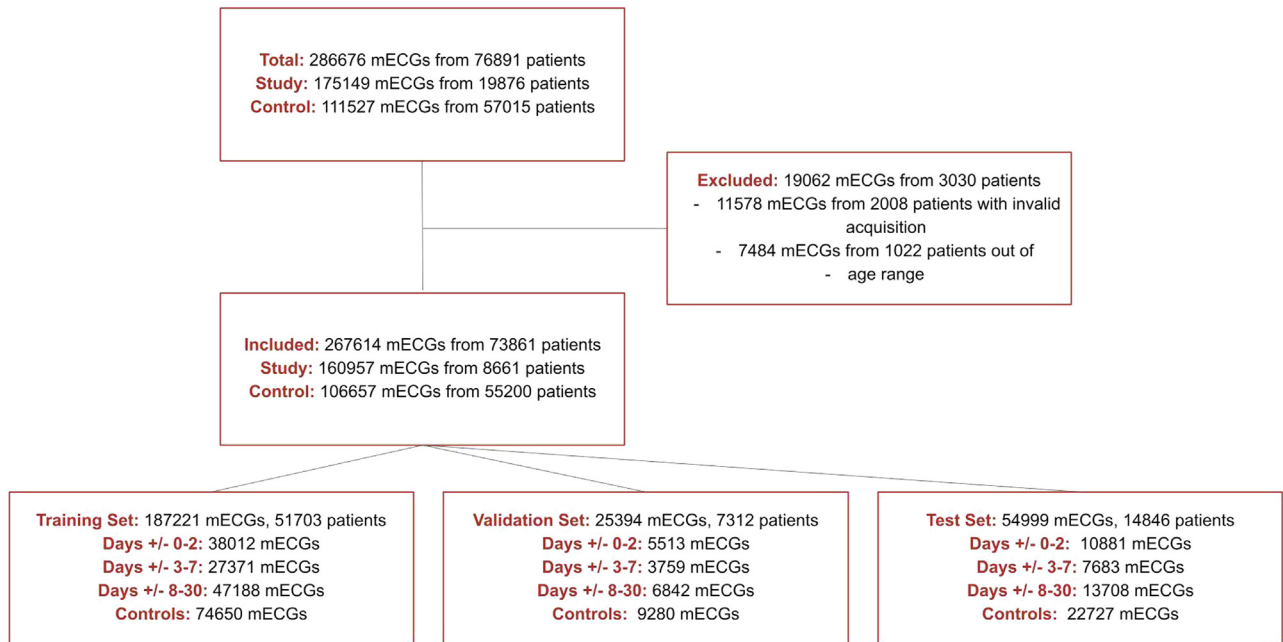
Figure 3   Diagram of the study data. mECG = mobile electrocardiogram.

preceding the AF event (BEFORE); and all mECGs collected in the days following the AF event (AFTER) (Figure 4). We evaluated the model on these subsets separately to ascertain whether AF events could be predicted prospectively or retro-spectively with higher fidelity (Figure 4). The number of users and mECGs in each study are shown in Supplemental Figures S1 and S2. Last, we examined model performance by age group (<65 years and ≥65 years) and sex.
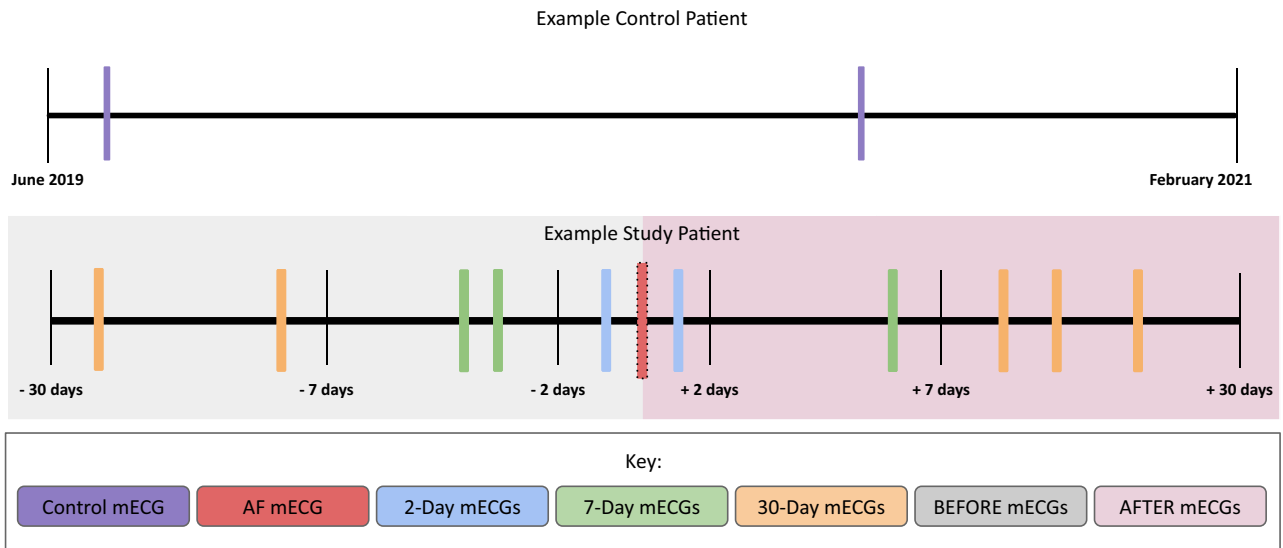

Figure 4   Diagram of the different test sets studied. **Top:** Mobile electrocardiograms (mECGs) from an example control user *(purple bars)*. Each control user assigned to the test set contributes a maximum of 2 mECGs, randomly sampled from the overall study window. These control mECGs are included in each version of the test sample, as they are not constrained by a central mECG. *Red bar* represents a randomly sampled atrial fibrillation (AF) mECG from a study user assigned to the test set. This is not included in the test set; however, it represents the origin of the ±30-day window for the user. **Bottom:** Distribution of mECGs for this user. *Blue bars* indicate mECGs in the 2-DAY test sample; *green bars* indicate mECGs in the 7-DAY sample; and *yellow bars* indicate mECGs in the 30-DAY sample. **Left** *(gray)* and **right** *(pink)* panels represent all study mECGs in the BEFORE test and AFTER tests, respectively. The AF mECG is not included in any of the test sets. *Blue, yellow, green,* and *purple bars* represent mECGs in the overall test set.

Table 1   Comparison of model performance by mECG screening period

Study	Sensitivity (95% CI)
2-DAY subsample	0.711 (0.709–0.713)
7-DAY subsample	0.708 (0.704–0.710)
30-DAY subsample	0.688 (0.685–0.690)
BEFORE	0.713 (0.710–0.716)
AFTER	0.695 (0.691–0.700)

Sensitivity and accuracy represent the same value for the subsample studies. Area under the receiver operating curve, specificity, and F1 not applicable for subsamples because performance is tested on study samples.
mECG = mobile electrocardiogram.

Statistical optimization

After each round of training, the fitted model was applied to the validation set. Carrying out inference on this intermediate set let us tune the internal parameters of the model. The final model was arrived at after such iterative training. Statistical optimization for each model included grid search over parameter spaces to find optimal hyperparameters: dropout equal to 0.1, Adamax optimizer, learning rate of 0.01, and a final dense layer of size 16. The final model was evaluated on the hold out test set using receiver operating characteristic, sensitivity, specificity, accuracy, and F1 score. For each secondary analysis described earlier, our model, architecture, and hyperparameters remained consistent, and our test set varied.

Results

Across the entire cohort, there were 73,861 included users with 267,614 mECGs. Overall, mean age was 58.14 ± 14.61 years; 25,991 users (35.1%) were female; and mean number of mECGs contributed per user was 3.62. Of these users, 18,661 individuals (25.3%) and 160,957 mECGs composed the study cohort. Mean age of the study cohort was 60.46 ± 14.55 years; 6098 users (32.7%) were female; and mean number of mECGs contributed was 8.62 ± 6.44. The control cohort included 55,200 users (74.7%) and 106,657 mECGs. Mean age of the control cohort was 54.64 ± 14.01 years; 19,893 (36.0%) were female; and mean number of mECGs contributed was 1.93 ± 0.25.

In total, 51,703 users and 187,221 mECGs were allocated to the training set, with 38,012 mECGs from the 2-DAY window, 27,371 mECGs from the 7-DAY window, 47,188 mECGs from the 30-DAY window, and 74,650 mECGs from the control sample. A total of 7,312 users and 25,394 mECGs were allocated to the validation set, with 5513 mECGs from the 2-DAY window, 3759 mECGs from the 7-DAY window, 6842 mECGs from the 30-DAY window, and 9280 mECGs from the control sample. A total of 14,846 users and 54,999 mECGs were allocated to the test set, with 10,881 mECGs from the 2-DAY window, 7683 mECGs from the 7-DAY window, 13,708 mECGs from the 30-DAY window, and 22,727 mECGs from the control sample. Users with paroxysmal AF contributed 60.15% of the mECGs in our study (Figure 3). Median number of

days between the first and second mECGs randomly sampled from each user in the control cohort was 238.4 (95% confidence interval [CI] 9.5–1314.1).

Our model achieved an overall area under the curve (AUC) score of 0.760 (95% CI 0.759–0.760), sensitivity of 0.703 (95% CI 0.700–0.705), specificity of 0.684 (95% CI 0.678–0.685), accuracy of 69.4% (95% CI 0.692–0.700), and F1 score of 0.694 (95% CI 0.694–0.700) on the test set comprising control samples and study samples from all 3 windows of interest (Table 1 and Figure 5). Prevalence, positive predictive value (PPV), and negative predictive value (NPV) were 0.601, 0.773, and 0.637 respectively, and, in practice, PPV and NPV would be impacted by real-life prevalence.

A comparison of model performance on study samples from the 2-DAY, 7-DAY, and 30-DAY windows of interest is given in Table 1. We report accuracy on each of the windows, which is equivalent to sensitivity for these subsamples. Generally, model performance was better on the 2-DAY window (sensitivity 0.711; 95% CI 0.709–0.713) and worse on the 30-DAY window (sensitivity 0.688; 95% CI 0.685–0.690), with performance on the 7-DAY window falling in between (sensitivity 0.708; 95% CI 0.704–0.710).

Secondary analyses

Performance characteristics of the model tested on sinus rhythm study mECGs stratified into before and after the AF event are given in Table 1. There were 14,268 BEFORE mECGs and 18,004 AFTER mECGs in the test set. The BEFORE and AFTER studies performed similarly to each other, with sensitivity of 0.713 (95% CI 0.709–0.717) and 0.695 (95% CI 0.692–0.699), respectively (Table 1).

When we examined the model performance by age group, the test set (including both control and study samples) contained 10,098 users (68.0%) <65 years old contributing 33,908 mECGs (32.0%) and 4748 users ≥65 years old contributing 21,091 mECGs.
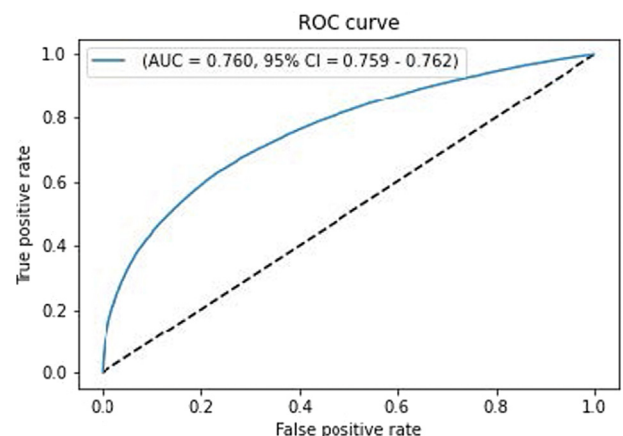

Figure 5   Model receiver operating characteristics (ROC) when trained with sinus rhythm mobile electrocardiograms obtained in all study windows. AUC = area under the curve; CI = confidence interval.

Table 2   Comparison of model performance measures by age and sex

Study	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	F1 (95% CI)
Test set	0.760 (0.758–0.760)	0.703 (0.700–0.705)	0.684 (0.678–0.685)	69.4% (69.2%–70.0%)	0.700 (0.694–0.700)
Age (y)					
18–64	0.743 (0.742–0.745)	0.615 (0.613–0.618)	0.750 (0.746–0.751)	0.680 (0.679–0.682)	0.679 (0.678–0.681)
64–99	0.749 (0.745–0.750)	0.807 (0.804–808)	0.508 (0.503–0.510)	0.720 (0.713–0.719)	0.715 (0.711–0.717)
Sex					
Male	0.765 (0.763–0.767)	0.714 (0.713–0.715)	0.674 (0.670–0.678)	70.0% (69.6%–70.0%)	0.700 (0.698–0.700)
Female	0.747 (0.746–0.751)	0.680 (0.678–0.686)	0.694 (0.691–0.696)	68.6% (68.4%–68.9%)	0.687 (0.685–0.691)

AUROC = area under the receiver operating curve.

The model performed marginally better in adults ≥ 65 years old (AUC 0.748; 95% CI 0.744–0.750) compared to adults <65 years old (AUC 0.743; 95% CI 0.742–0.745) (Table 2).

When we examined the model performance by sex, the test set (including both control and study samples) contained 10,252 male users (69.0%) contributing 36,436 mECGs and 4594 (31.0%) female users (31.0%) contributing 18,563 mECGs. Model performance was similar for men (AUC 0.765; 95% CI 0.763–0.767) and women (AUC 0.747; 95% CI 0.746–0.751) (Table 2).

Discussion
In our study of more than 260,000 mECGs in 73,861 Alive-Cor 6L users, we found that neural networks were able to accurately predict AF events from sinus rhythm mECGs with good performance metrics. We demonstrated comparable model performance on data obtained before or after the AF event, which suggests AF events can be predicted before they occur. Additionally, we showed that the accuracy of AF prediction improved when mECGs were screened closer to the AF event, with the predictions on the 2-DAY window outperforming the 7-DAY and 30-DAY windows. Taken together, our results suggest AI-enabled AF event prediction using mECGs is time-dependent because model performance is best when trained with mECGs in the immediate temporal vicinity of an AF event but diminishes with time. These results highlight the utility and feasibility of using AI techniques and mECG data to predict AF events, which may be a valuable screening tool for reducing morbidity and mortality of patients with AF.

To our knowledge, this study is the first to use AI techniques to predict AF events prospectively and retrospectively using sinus rhythm mECG data and to analyze a temporal relationship with model performance. There are several potential advantages to leveraging mECG data vs 12-lead data to predict AF using AI, including more frequent ECG sampling, ease of acquisition, and administration without the need to travel to a health care facility, ultimately improving access to AF screening for those without regular medical care. In addition, mECGs can be uploaded to a central cloud-based data repository, enabling multiple institutions to implement AI-based AF event prediction paradigms, whereas 12-lead ECGs must be processed at the institutional level and require significant technical support to abstract from medical records. Thus, mECG-based AI techniques represent a significantly more scalable, wide-reaching, and cost-effective means of predicting AF.

Our results can be taken in the context of medical screening tests including B-type natriuretic peptide for heart failure (AUC 0.60–0.70), Papanicolaou smear for cervical cancer (AUC 0.70), and CHA_2DS_2-VASc Score for stroke risk (AUC 0.57–0.72) as compared by Attia et al.[7] Although we demonstrated robust predictive abilities when applied to mECG data, our model did not outperform AI prediction of AF when applied to 12-lead ECGs. In this context, Attia et al demonstrated an AUC of 0.90, and Raghunath demonstrated an AUC of 0.85.[7,8] There may be several reasons for the discrepancies between their findings and ours. The number of mECGs and users included in our dataset was 4 times smaller than in the study by Attia et al and an order of magnitude smaller compared to the work by Raghunath. As more mECG data are collected, we would expect higher training capabilities and improved model performance. Additionally, the mECG sampling rate and spatial resolution are 40% and 50% lower, respectively, than traditional 12-lead ECGs (sampling rate is 300 Hz in mECGs and 500 Hz in traditional 12-lead ECGs; mECGs do not have leads in the frontal axis). Because neural networks recognize hidden patterns from minute details, performance likely would decrease as resolution decreases. The tradeoff between better spatial resolution vs ease of implementation, scalability, and widespread adoption will have to be considered when performing AI-based analyses on ECG data. Finally, although mECG devices are gaining popularity, we are using data from Alivecor users who already are monitoring health. Therefore, it is likely that some patients who appear in the control actually have paroxysmal AF, making model prediction more challenging, mimicking a real-world use of such a tool.

We found that AF prediction was most accurate when our algorithm was tested on mECGs closest to the AF event. One explanation for this finding is an increase in nonsinus electrical activity occurring in close temporal proximity to AF events. It has been hypothesized that subtle, nonsinus electrical activity that subtly alters ECG morphology may immediately precede or follow AF, and these minute changes would facilitate AI-based detection of AF.[17,18] Our findings seem to confirm this hypothesis, as performance of the model on 2-DAY data was superior to that of the model on 7-DAY and

30-DAY data, respectively. This finding carries important AI training implications: to best predict AF development, AI algorithms should be trained on ECG data as close as possible to an AF event.

The ability to predict AF from mECGs may have a significant clinical impact. Because paroxysmal AF can be asymptomatic, thromboembolic stroke often occurs before a diagnosis of AF is made.[19] By identifying from mECG data those individuals at high risk for AF, clinicians may be able to selectively implement more aggressive surveillance measures, such as prolonged outpatient monitoring, more frequent mECG sampling, or prescription of an implantable loop recorder, so that anticoagulation can be prescribed as soon as AF is diagnosed but before a stroke occurs. The ability to identify those at high risk for AF with AI-enabled algorithms thus may result in more cost-effective and efficient use of traditional long-term ECG monitoring techniques. In addition to primary prevention of stroke, neural networks may prevent additional strokes.

Predicting future AF events may enable novel paradigms of how pharmacotherapy can be used for those with AF. For example, rhythm-controlling medications are sometimes used to maintain a sinus rhythm in highly symptomatic individuals with paroxysmal AF, but these drugs are associated with significant and potentially dangerous side effects if taken for the long term. If a future AF event could be predicted with high accuracy, these medications potentially could be taken as needed before an AF event occurs, which may be useful in reducing side effects, decreasing pill burden, and increasing medication adherence. Similarly, in the case of anticoagulation medication for stroke prevention, AI-based event prediction algorithms could be leveraged to tailor anticoagulation duration to periods of high likelihood of an AF event. This has the potential to reduce the frequency of adverse bleeding events and decrease pill burden for the primary prevention of stroke, resulting in more patient-centered and effective clinical care.

Study limitations
Our study cohort is small compared to studies of AI-enable AF prediction in 12-lead ECGs, which likely resulted in lower robustness with our neural networks. However, this is the first and largest study using mECGs for such prediction models, and as mECGs become more widely used, the predictive power of neural networks likely will increase. Second, the spatial resolution of mECGs is less than that of 12-lead ECGs, possibly resulting in the neural network missing detail in ECG changes predictive of AF. The tradeoff between better spatial resolution and ease of implementation, scalability, and widespread adoption will have to be considered when performing AI-based analysis on mECG data. Third, the Alivecor AF detection package does not distinguish between AF and atrial flutter, an atrial arrhythmia with a different mechanism than AF. However, because atrial flutter also carries an increased risk of stroke,[20] AI models that predict both of these rhythms would be useful. Fourth,

we do not have characteristic information (eg, additional clinical indicators, socioeconomic standing, or racial background) for our study population, limiting the applicability of our study to real-life clinical settings. Finally, because our results included only participants using the Alivecor 6L device, they may not be generalizable to users of other mECG devices. Further studies assessing the predictive capabilities of AI in other mECG platforms, with demographic context, may confirm our results.

Conclusion
We demonstrated that neural networks can predict with good performance the development of AF using sinus rhythm mECG data. As the screening approached an AF event, we observed that the predictive power of the CNN increased. Ultimately, mECG data may lower barriers to implementing AI-based AF event prediction systems in the modern health care setting because of its scalability, availability, and cost-effectiveness.

Data Availability
All reasonable requests for raw and analyzed data and related materials will be reviewed for patient privacy or confidentiality constraints. The code for the models is available upon request.

Funding Sources

Disclosures
Drs Schram and Albert are employees of AliveCor, Inc., which produces the portable 6-lead ECG machines. They were instrumental in obtaining and processing the data extracted from the Alivecor Data repository. They were not involved with building the neural network for the current study, and their conflict did not influence the results of the study. All other authors declare no competing interests, conflicts of interest, or relevant financial disclosures.

Appendix
Supplementary data
Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.cvdhj.2023.01.002

References
1. Iwasaki Y, Nishida K, Kato T, Nattel S. Atrial fibrillation pathophysiology. Circulation 2011;124:2264–2274.
2. Hart RG, Benavente O, McBride R, Pearce LA. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. Ann Intern Med 1999;131:492–501.

3. Anter E, Jessup M, Callans DJ. Atrial fibrillation and heart failure. Circulation 2009;119:2516–2525.

4. Dilaveris PE, Kennedy HL. Silent atrial fibrillation: epidemiology, diagnosis, and clinical impact. Clin Cardiol 2017;40:413–418.

5. Seet RCS, Friedman PA, Rabinstein AA. Prolonged rhythm monitoring for the detection of occult paroxysmal atrial fibrillation in ischemic stroke of unknown cause. Circulation 2011;124:477–486.

6. Kishore A, Vail A, Majid A, et al. Detection of atrial fibrillation after ischemic stroke or transient ischemic attack. Stroke 2014;45:520–526.

7. Attia ZI, Noseworthy PA, Lopez-Jimenez F, 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. Lancet 2021;394:861–867.

8. Raghunath S, Pfeifer JM, Ulloa-Cerna AE, et al. Deep neural networks can predict new-onset atrial fibrillation from the 12-lead ECG and help identify those at risk of atrial fibrillation–related stroke. Circulation 2021;143:1287–1298.

9. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. Stroke 1991;22:983–988.

10. Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019;381:1909–1917.

11. Hickey KT, Hauser NR, Valente LE, et al. A single-center randomized, controlled trial investigating the efficacy of a mHealth ECG technology intervention to improve the detection of atrial fibrillation: the iHEART study protocol. BMC Cardiovasc Disord 2016;16:152.

12. Marcus D, Vivien N, Noé B, et al. The WATCH AF Trial: SmartWATCHes for Detection of Atrial Fibrillation. JACC Clin Electrophysiol 2019;5:199–208.

13. Lopez Perales CR, Van Spall HGC, Maeda S, et al. Mobile health applications for the detection of atrial fibrillation: a systematic review. Europace 2021;23:11–28.

14. Association for the Advancement of Medical Instrumentation. ANSI/AAMI/IEC 60601-2-47:2012 (R2016). 2012.

15. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Conf Comput Vis Pattern Recognit 2016;770–778.

16. US Food and Drug Administration. FDA Kardia AI Clearance. https://www.accessdata.fda.gov/cdrh_docs/pdf18/K181823.pdf. Accessed December 12, 2021.

17. Warraich HJ, Gandhavadi M, Manning WJ. Mechanical discordance of the left atrium and appendage: a novel mechanism of stroke in paroxysmal atrial fibrillation. Stroke 2014;45:1481–1484.

18. Bellotti P, Spirito P, Lupi G, Vecchio C. Left atrial appendage function assessed by transesophageal echocardiography before and on the day after elective cardioversion for nonvalvular atrial fibrillation. Am J Cardiol 1998;81:1199–1202.

19. Lin H-J, Wolf PA, Benjamin EJ, Belanger AJ, D'Agostino RB. Newly diagnosed atrial fibrillation and acute stroke. Stroke 1995;26:1527–1530.

20. Vadmann H, Nielsen PB, Hjortshøj SP, et al. Atrial flutter and thromboembolic risk: a systematic review. Heart 2015;101:1446–1455.