Letter to the Editor



# Machine learning in atrial fibrillation—racial bias and a call for caution

**Hiten Doshi[1], Jay Chudow[2], Kevin Ferrick[2], Andrew Krumerman[2]**

[1]Tufts Medical Center, Boston, MA, USA; [2]Albert Einstein College of Medicine, Montefiore Medical Center, Bronx, NY, USA

*Correspondence to:* Andrew Krumerman, MD. Department of Medicine, Division of Cardiology, Division of Cardiac Electrophysiology, Montefiore Medical Center – Albert Einstein College of Medicine, 111 East 210th Street, Bronx, NY 10467-2401, USA. Email: akrumerm@montefiore.org.

Received: 19 May 2021; Accepted: 31 August 2021; Published: 30 September 2021.

doi: 10.21037/jmai-21-12

**View this article at:** https://dx.doi.org/10.21037/jmai-21-12

Early diagnosis of atrial fibrillation (AF), a common arrhythmia that can cause adverse events such as stroke, is a major clinical challenge. Due to its often asymptomatic and paroxysmal nature, AF is easily missed on single electrocardiograms (ECGs), making outpatient screening challenging. As a result, patients may not receive a timely diagnosis, with up to 5% of all AF cases being diagnosed at the time of stroke (1). Various machine learning (ML) models, primarily involving supervised ML methods, have been developed with the hopes of bringing an effective population screening tool to the forefront. While these models show strong performance in their respective studies, data regarding their effectiveness across racial groups is lacking. Therefore, using ML for AF screening requires two important considerations: (I) any biases in the training set data will be perpetuated in the predictions that the models offer; (II) AF has a known racial paradox, where traditional risk factors that were derived from a largely Caucasian population have a weaker correlation with AF incidence in Black patients. Below, we elaborate on these points and argue that while ML presents a unique opportunity to increase the detection of AF, it also deserves special caution to avoid reinforcing existing healthcare disparities.

ML AF screening tools are commonly developed using ECG data about p-waves, R-R intervals, heart rate, and other parameters. While this has shown the ability to produce strong predictive models, the actual data sources deserve scrutiny (2). A recently published systematic review identified that while more than 100 publications exist using ECG data to develop ML models, more than half of them used the same four open-access ECG databases (3). In theory, this is not necessarily problematic, and it is understandable that so many studies reuse well known and freely available datasets. Ideally, however, the datasets would report a sufficient level of patient diversity to well represent the entire US population. Instead, many of the most commonly used ECG datasets only report limited demographic data, including the patient's age, gender, and/or baseline clinical characteristics, without reporting racial or ethnic background. Considering the known racial differences that exist in several baseline ECG parameters, including left ventricular hypertrophy, right axis deviation, bundle branch blocks, and others, transparency about racial demographic information in these datasets is critical (4). *Table 1* summarizes the most commonly used ECG databases, as well as the readily available demographic information provided by each.

The reuse of these datasets carries particular concern in the diagnosis of AF, a disease with a known "racial paradox". This paradox refers to the fact that while Black patients have a higher burden of AF risk factors including hypertension, diabetes, congestive heart failure, and others, they paradoxically have a lower incidence of AF (5). Many explanations for this paradox have been proposed, including underdiagnosis of AF in Black patients due to lower healthcare access, regional genetic variations, or an unequal influence of certain risk factors between racial groups (6-8). In either case, the presence of this paradox makes data transparency in AF an even greater priority. In the same way that traditional risk factors for AF showed worse correlations with incidence in Black patients, we may now be developing ML models with the same shortcomings.

One solution is for hospital systems to develop AF models using their own internal databases. The Mayo

**Table 1** Demographic data provided by ECG databases

| Dataset number | Name of database | # Records | Age reported (y/n) | Sex reported (y/n) | Baseline clinical characteristics reported (y/n) | Racial and/or ethnic data reported (y/n) | Open Access |
|---|---|---|---|---|---|---|---|
| 1 | MIT-BIH Arrhythmia Database | 47 | x | x | – | – | x |
| 2 | PhysioNet Computing in Cardiology Challenge 2017 | 8,582 | – | – | – | – | x |
| 3 | PTB Diagnostic ECG Database | 549 | x | x | x | – | x |
| 4 | MIT-BIH Atrial Fibrillation Database | 25 | – | – | x | – | x |
| 5 | Mayo Clinic Digital Data Vault | 649,931 | x | x | – | – | – |

Datasets 1–4 are the most commonly used open access ECG databases, and can be found on PhysioNet, a community resource developed under the auspices of the National Institute of Health. X, yes; –, No.

Clinic, for example, used its own digital data library to develop a ML model that identifies patients with AF from sinus rhythm ECGs (9). Although the racial demographics of this dataset were also not reported, thereby limiting the model's utility in external populations, their initiative is still a step in the right direction. The use of a large and internally derived dataset maximizes the chance that the training data will appropriately reflect the Mayo Clinic's patient population. While not every health system has the resources to develop their own high quality ML models, the increasing embrace of electronic medical records, wearable devices, and other sources of patient data will help make data and ML more accessible. Taking deliberate actions now, prior to the widespread implementation of ML in routine clinical practice for AF, will lead us towards resolving rather than reinforcing health disparities. One possible solution would be to include demographic data as variables when training models, but this can be problematic. Allowing algorithms to consider race, ethnicity, or other demographic data when offering predictions is risky, especially in ML where the interactions between variables and outcomes are often not transparent. With this in mind, we have the following recommendations: (I) ML models developed for AF screening should ensure the use of diverse training sets; (II) descriptions of datasets should ensure the reporting of racial and ethnic information; (III) health systems should validate ML models in their own populations prior to implementation in clinical practice for AF. A strong commitment to these principles will make ML a promising tool to increase AF detection rates and promote early diagnosis and treatment for all patients.

## Acknowledgments

*Funding:* None.

## Footnote

*Provenance and Peer Review:* This article was a standard submission to the journal. The article has undergone external peer review.

*Peer Review File:* Available at https://dx.doi.org/10.21037/jmai-21-12

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/jmai-21-12). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## References


1. Lubitz SA, Yin X, McManus DD, et al. Stroke as the Initial Manifestation of Atrial Fibrillation: The Framingham Heart Study. Stroke 2017;48:490-2.
2. Mincholé A, Camps J, Lyon A, et al. Machine learning in the electrocardiogram. J Electrocardiol 2019;57S:S61-4.
3. Hong S, Zhou Y, Shang J, et al. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. Comput Biol Med 2020;122:103801.
4. Friedman A, Chudow J, Merritt Z, et al. Electrocardiogram abnormalities in older individuals by race and ethnicity. J Electrocardiol 2020;63:91-3.
5. Alonso A, Agarwal SK, Soliman EZ, et al. Incidence of atrial fibrillation in whites and African-Americans: the Atherosclerosis Risk in Communities (ARIC) study. Am Heart J 2009;158:111-7.
6. Gbadebo TD, Okafor H, Darbar D. Differential impact of race and risk factors on incidence of atrial fibrillation. Am Heart J 2011;162:31-7.
7. Shulman E, Chudow JJ, Shah T, et al. Relation of Body Mass Index to Development of Atrial Fibrillation in Hispanics, Blacks, and Non-Hispanic Whites. Am J Cardiol 2018;121:1177-81.
8. Shulman E, Chudow JJ, Essien UR, et al. Relative contribution of modifiable risk factors for incident atrial fibrillation in Hispanics, African Americans and non-Hispanic Whites. Int J Cardiol 2019;275:89-94.
9. 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 2019;394:861-7.


doi: 10.21037/jmai-21-12
**Cite this article as:** Doshi H, Chudow J, Ferrick K, Krumerman A. Machine learning in atrial fibrillation—racial bias and a call for caution. J Med Artif Intell 2021;4:6.