Check for updates

Prediction of Atrial Fibrillation Using Machine Learning: A Review

Andrew S. Tseng and Peter A. Noseworthy*

Department of Cardiovascular Diseases, Mayo Clinic, Rochester, MN, United States

There has been recent immense interest in the use of machine learning techniques in the prediction and screening of atrial fibrillation, a common rhythm disorder present with significant clinical implications primarily related to the risk of ischemic cerebrovascular events and heart failure. Prior to the advent of the application of artificial intelligence in clinical medicine, previous studies have enumerated multiple clinical risk factors that can predict the development of atrial fibrillation. These clinical parameters include previous diagnoses, laboratory data (e.g., cardiac and inflammatory biomarkers, etc.), imaging data (e.g., cardiac computed tomography, cardiac magnetic resonance imaging, echocardiography, etc.), and electrophysiological data. These data are readily available in the electronic health record and can be automatically queried by artificial intelligence algorithms. With the modern computational capabilities afforded by technological advancements in computing and artificial intelligence, we present the current state of machine learning methodologies in the prediction and screening of atrial fibrillation as well as the implications and future direction of this rapidly evolving field.

Keywords: atrial fibrillation, electrocardiogram, echocardiography, risk factor, prediction, deep learning, machine learning

INTRODUCTION

Atrial fibrillation (AF) is the most common arrhythmia worldwide with its burden expected to continue to increase with the aging population. AF is diagnosed clinically, requiring detection of the arrhythmia on formal electrocardiographic testing. Improvements in monitoring technology, including high-fidelity long-term monitors, have increased the yield for the detection of AF, thereby enhancing our knowledge of the true clinical burden of AF.

Beyond detection, there has been immense interest in prediction of AF using both clinical risk factors as well as objective testing. Numerous clinical risk scores have been proposed, incorporating readily available variables from the patient's medical history, such as age, ethnicity, height, weight, blood pressure, smoking status, medication use, and comorbidities (Schnabel et al., 2009; Chamberlain et al., 2011; Alonso et al., 2013; Suenari et al., 2017; Li et al., 2019; Hu and Lin, 2020; Lip et al., 2020; Himmelreich et al., 2021). Abnormalities in both cardiac and inflammatory biomarkers have been shown to augment the predictive ability of clinical prediction scores (O'Neal et al., 2016). Structural cardiac abnormalities, including atrial fibrosis and atrial enlargement, as well as associated manifestations on physiologic parameters such as mitral inflow Doppler and atrial strain have been shown to be predictive of AF (De Vos et al., 2009; Caputo and Mondillo, 2012; Hwang et al., 2015). Likewise, electrocardiographic (ECG), particularly P wave morphology, has been well-studied and shown to have predictive utility. Overall, there is an abundance of clinical variables that have been shown to be predictive of AF, individually or in limited pairings.

With advancements in artificial intelligence technology and the rapid accumulation of digital clinical data, machine learning has the potential to analyze and synthesize seemingly disparate variables to predict AF in such a way that vastly surpasses conventional methods (Siontis et al., 2020). Machine learning algorithms can not only assist in processing imaging or electrocardiographic data, but it may also be able to incorporate and interpret large amounts of clinical data and discover new clinical patterns and concepts. We seek to present the latest review of conventional and machine learning methodologies in the prediction of AF.

THE PRESENT STATE OF MACHINE LEARNING TECHNIQUES

At the core of machine learning is the convergence of statistical analytics and computer engineering. Machine learning algorithms are able to process complex inputs, such as images, and discern subtle relationships that may not be evident with traditional statistical methods. Machine learning techniques can be categorized broadly into three categories: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning requires labels during training, such as the presence or absence of incident AF. Therefore, the algorithm is provided with both the input variables as well as outcome labels. Unsupervised learning seeks to identify relationships within the data without the assistance of labels. Various methods such as clustering have been described for this method of learning. Reinforcement learning uses the concept of reward maximization, in which the machine learning algorithm assumes the role of an agent that receives either positive or negative reinforcement to guide decision making (Thrun and Littman, 2000; Koohy, 2017; Géron, 2019). For the purposes of this review, we will discuss the most used learning method, supervised learning.

Supervised learning itself utilizes different methods including regression modeling, random forests, and neural networks. In regression modeling, both with and without machine learning, preselected variables undergo regression analysis to determine their ability to predict an outcome. Machine learning improves upon these traditional modeling techniques by its ability to analyze large and complex datasets. Techniques include classification algorithms such as Support Vector Machine and K-Nearest Neighbor (Sultana et al., 2016). Random forests utilize branching decision trees, empirically deriving thresholds to determine how the data should be split (Koohy, 2017; Géron, 2019; Uddin et al., 2019).

Neural networks have fundamentally changed the machine learning landscape. Fundamentally, the network architecture is comprised of layers and processing units within each layer called nodes. Data is analyzed in one layer and then transmitted to the next layer, such that a node in a deeper layer receives inputs from one or more nodes in the prior layer. All neural networks have an input layer to process input data and an output layer while a deep neural network continues numerous "hidden" intermediary layers and nodes. Convolutional neural networks utilize the concept of "convolutions," whereby nodes in a deeper layer only receive input from select subset of nodes from the previous layer. Therefore, these networks seek to identify local correlations and preserve local special dependences, which is particularly important for image processing. It also allows for more efficient computational processing by reducing the input data into smaller localized convolved features via methods of dimensionality reduction (Stankovic and Mandic, 2021).

There are a vast array of different machine learning techniques, which by themselves can be the subject of reviews and textbooks. For the clinician, we have summarized the different supervised machine learning techniques, including names of techniques one might encounter, as well as the general advantages and disadvantages in **Table 1**.

FROM CLINICAL DATA

Validated clinical risk scores to predict AF, such as the FHS, ARIC, CHARGE-AF, C2HEST, and HATCH score, utilize readily obtainable clinical variables, such as age, ethnicity, height, weight, blood pressure, smoking status, antihypertensive medication use, history of diabetes, heart failure myocardial infarction, etc. Based on these readily available variables from the patient history, these risk scores have shown adequate model discrimination for the prediction of incident AF (area under the receiver operator curve, AUCs, generally around 0.70) (Schnabel et al., 2009; Chamberlain et al., 2011; Alonso et al., 2013; Suenari et al., 2017; Li et al., 2019; Hu and Lin, 2020; Lip et al., 2020; Himmelreich et al., 2021). AUCs, or c-statistic, are commonly used in studies of diagnostic test performance as an overall indicator of test performance (Bradley, 1997). Other measures of test performance, though not universally reported, include accuracy (proportion of correct assessments), precision (or positive predictive value), and recall (or sensitivity). Due to inconsistencies with reporting these other measures of test performance, which limits comparison among studies, we will largely focus on AUCs. The studies for these validated clinical risk scores to predict AF are summarized in **Table 2**.

The addition of serologic testing of common cardiac biomarkers, including natriuretic peptides and C-reactive protein, has been shown to enhance the predictive ability of such clinical risk scores (Sinner et al., 2014). Additional markers of chronic kidney disease, such as Cystatin C, and endothelial dysfunction have also been shown to be associated with AF, though no studies have been shown that the addition of these parameters enhances the predictive ability of existing clinical risk scores (O'Neal et al., 2016).

With the abundance of clinical and laboratory data available in digital format, recent investigators have started to evaluate the use of machine learning in predicting AF using the electronic health record. To facilitate this, organizations have developed a common data models for analysis, one prime example being the Observational Medical Outcomes Partnership Common Data Model in an effort to synchronize data from disparate sources for systematic analysis (FitzHenry et al., 2015). In a recent large study of nearly 2 million patients from the University of Colorado health systems by Tiwari et al. (2020) investigators applied a

TABLE 1 | Advantages and disadvantages of different supervised machine learning techniques.

Examples	Advantages	Disadvantages
Linear/logistic regression	Simple and easy to implement	Reduced accuracy with variables with complex relationships
k-Nearest Neighbor	Simple and can handle noisy instances or instances with missing attribute value	Computationally taxing, does not identify variables that are important for classification
Support vector machines	More robust than ordinary regression and can classify semi-structured/unstructured data	Computationally taxing, does not handle data noise well
Random forests	Performs well in large datasets and identifies variables that are important for classification	Computationally taxing, easy to overfit
Neural networks	Detects complex non-linear relationships between variables	Computationally taxing, cannot access decision making process ("blackbox")

TABLE 2 | Original and validation studies of clinical AF risk scores.

Clinical AF risk score	Original study	Example validation*
FHS (Age, sex, BMI, SBP, hypertension treatment, PR interval, clinically significant cardiac murmur, CHF)	4,764 patients from the United States AUC 0.78 (95% CI 0.76–0.80) (Schnabel et al., 2009)	49,599 patients from the United States 0.734 (0.724–0.744) (Shulman et al., 2016)
ARIC (Age, race, height, SBP, hypertension treatment, smoking status, precordial murmur, left ventricular hypertrophy, left atrial enlargement, DM, CAD, CHF)	14,546 patients from the United States AUC 0.765; 95% CI, 0.748–0.781 (Chamberlain et al., 2011)	None
CHARGE-AF (Age, ethnicity, height, weight, blood pressure, smoking, antihypertensive use, DM, CHF, MI)	18,556 patients from the United States AUC 0.765 (95% CI: 0.748–0.781) (Alonso et al., 2013)	114,475 patients from the Netherlands AUC 0.74 (95% CI: 0.73–0.74) (Himmelreich et al., 2021)
C2HEST (CAD/COPD, hypertension, age, CHF, hyperthyroidism)	471,446 patients from China AUC 0.75 (95% CI: 0.73–0.77) (Li et al., 2019)	1,047,330 patients from Denmark AUC 0.588 (95% CI: 0.585–0.591) (Lip et al., 2020)
HATCH (Hypertension, age, stroke/TIA, COPD, CHF)	670,804 patients from Taiwan AUC 0.716 (95% CI: 0.710–0.723) (Suenari et al., 2017)	692,691 patients from Taiwan AUC 0.771 (no CI provided) (Hu and Lin, 2020)

AF, atrial fibrillation; AUC, area under the curve; BMI, body mass index; CAD, coronary artery disease; CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; DM, diabetes mellitus; MI, myocardial infraction; SBP, systolic blood pressure; TIA, transient ischemic attack.
*Studies selected for explanatory purposes and may not be an exhaustive list.

machine learning model to over 200 most common health record features, including demographics and comorbidity data, and derived a model with an AUC of 0.79 to detect incident AF in a 6 month timeframe, which is in line with non-machine learning clinical AF risk scores. In another study of over 2 million primary care patients from the United Kingdom by Sekelj et al. (2021) another machine learning algorithm achieved an AUC of 0.83 in the development dataset and 0.87 in the validation dataset to detect incident AF in a registry that spanned 7 years, indicating better performance compared to traditional risk scores.

When comparing the AI algorithms to the traditional risk scores, many factors may impact and limit the interpretation of the test performance. Firstly, there is significant variation in the duration of follow-up for each study, ranging from as short as 6 months to more than 10 years. This clearly significantly impacted the proportion of patients at study termination with incident AF (1% vs. 10%, respectively) (Schnabel et al., 2009; Tiwari et al., 2020). It is possible that limited follow-up such as 6 months in the Tiwari AI study may have reduced the test performance in part due to the limited duration of follow-up,

where "false positives" (i.e., AI screening positive, AF negative at 6 months) would have been "true positives" if given sufficient time to manifest or vice versa with "true negatives" at the end of study turning into "false negatives" (Tiwari et al., 2020).

In a recent study by Hill et al. (2019) of nearly 3 million patients in the United Kingdom, the investigators compared a machine learning algorithm with time-varying covariates to the CHARGE-AF risk score. The use of time-varying covariates represents yet another technique in neural networks, in which the input covariates are not static but are allowed to be incorporated into the model at varying time points during the study period. This means that the temporal association between a covariate and the outcome becomes another critical factor during the development of these neural networks. In this study, the found that the time-varying model had an AUC of 0.827 while the traditional CHARGE-AF risk score applied to the same population had an AUC of 0.725. Using the time-varying methodology, they were able to determine that congestive heart failure diagnosed within the most recent 91-day quarter contributed the most to the prediction of incident AF. This study

not only showed the benefits of using different machine learning techniques to extract potentially clinically relevant predictors (such as time-dependent variables), but also that the machine learning algorithms performed more robustly than traditional risk scores (Hill et al., 2019). While these algorithms have not been tested prospectively nor have they been validated in external health systems, the size and scale of these massive projects far exceed previous studies of conventional clinical risk scores for AF, and shows the increasing promise of utilizing easily accessible data from the electronic health record to predict the risk of incident AF.

FROM CARDIAC IMAGING DATA

AF is often associated with distinct structural heart abnormalities that are apparent on cardiac imaging, including echocardiography, CT and MRI. Oftentimes, these structural abnormalities result from conditions that predispose patients to AF, such as diastolic dysfunction, but AF can also itself lead to valvular regurgitation. From an echocardiography perspective, previous studies have shown that left atrial volumes, measures of diastolic dysfunction, ventricular wall thickness, strain echocardiography can be associated with the risk of new-onset AF (Xu et al., 2011; Caputo and Mondillo, 2012; Hirose et al., 2012). Newer, non-conventional measurements such as the total atrial conduction time, as a marker of atriopathy, was also shown to be associated with development of AF in a smaller cohort of 249 patients (De Vos et al., 2009). Cardiac CT to evaluate the left atrial appendage have demonstrated mixed results on prediction of AF after AF ablation (Ebersberger et al., 2020). However, left atrial thickness as a marker of atriopathy on cardiac CT has been shown to be associated with increased risk of transition for paroxysmal AF to chronic AF as well as low-voltage areas as potential sites of ablation (Nakamura et al., 2011; Nakatani et al., 2020). Given the unique ability of MRI to evaluate tissue characteristics, left atrial fibrosis by late gadolinium enhancement on cardiac MRI has been shown to be associated with new-onset AF. One study with 182 patients evaluated the predictive ability of left atrial fibrosis>6% and derived an AUC of 0.67, which was further enhanced to 0.80 after adding history of hypertension and left ventricular ejection fraction (Siebermair et al., 2019). Overall, the use of these imaging parameters to predict AF have largely been restricted to small association and procedural studies, and there has not been systematic use of imaging data to develop or refine existing risk scores for predicting AF.

Machine learning has likewise begun to make headway in image analysis. Unlike the categorical or numerical input of data from the electronic health records, images require additional sophisticated methodologies when applying machine learning, yet the fundamental theory remains similar (Fu et al., 2019). Small-scale studies have started to investigate the use of machine learning on cardiac imaging. In a small study on cardiac CT using machine learning to evaluate left atrial and pulmonary vein morphology in 203 patients undergoing AF ablation, the machine learning algorithm was able to predict AF recurrence after ablation using these CT images with an AUC of overall AUC of 0.87 (Firouznia et al., 2021). A similar study of 68 patients using cardiac CT left atrial morphology to predict AF recurrence after ablation demonstrated an AUC of 0.78 when combining imaging and clinical features (Atta-Fosu et al., 2021).

However, there have not been investigations in the use of machine learning in cardiac imaging to predict new-onset AF. Given multiple factors, including the complexity of image processing, machine learning in cardiac imaging has focused on image acquisition, processing, and basic interpretation (Chang et al., 2020). Future studies will be needed to develop the role of machine learning in prognosis and detection of non-imaging diagnoses such as AF. As such, large population-based studies may not be feasible, related to the costs of screening asymptomatic patients with imaging and significant selection bias for patients who have indications to undergo cardiac imaging tests. Nonetheless, machine learning in cardiac imaging for AF will undoubtedly play an important role in periprocedural prognosis and management, and perhaps with well-designed studies can help with the prediction of AF.

FROM ELECTROPHYSIOLOGICAL DATA

Pathophysiologic changes in AF can also manifest itself as abnormalities on electrophysiology testing, such as electrocardiography and invasive intracardiac electrograms. Previous studies have shown that ECG findings, such as P-wave duration, dispersion and amplitude as well as premature atrial contraction morphology and frequency, have been shown to be predictive of incident AF, achieving AUCs ranging from 0.69 to 0.87 (Dilaveris et al., 2000; Thong et al., 2004; Yoshizawa et al., 2014). One study evaluated premature atrial contraction characteristics and percent burden as a risk factor for AF among 652 patients who underwent Holter monitoring, with an AUC of 0.58 (Im et al., 2018).

Intracardiac electrograms are generally obtained during an electrophysiology study in patients with known or suspected arrhythmia. Therefore, there have not been studies specifically evaluating the predictive ability of electrogram features on new-onset AF. However, elements of the intracardiac electrogram have been shown to be correlated with the risk of AF recurrence after ablation. For example, in one study on 140 patients, multiple electrogram characteristics including dominant frequency, regularity index and organizational index of fibrillatory electrograms have shown predictive value for AF recurrence after AF ablation (Szilagyi et al., 2018).

The use of machine learning on the ECG to predict new-onset AF has been the subject of immense inquiry recently. Unlike the use of machine learning to process cardiac imaging, the processing of electrocardiographic signals is highly standardized using a static time series dataset and more easily interpretable compared to a series of images, including moving images like in echocardiography. In a large study from the Mayo Clinic of over

600,000 ECGs in normal sinus rhythm, a convolutional neural network was developed with a robust AUC of 0.87 for predicting new-onset AF, with further augmentation of the AUC to 0.90 for patients with multiple ECGs (Attia et al., 2019). A small study by Ebrahimzadeh et al. (2018) in 53 patients of extended ECG recordings sought to evaluate different machine learning techniques using heart rate variability analysis in extended ECG monitoring to predict initiation of AF. In this self-controlled study, all patients had an episode of paroxysmal AF, in which a 5-min ECG segment obtained 30 min prior to the onset of AF ("AF" label) was compared to a 5-min ECG segment obtained 45 min after termination of AF ("non-AF" label). Unlike the convolutional neural network used in the Mayo Clinic study in which the features of the neural network are hidden, the investigators identified multiple features from the heart rate variability signal, including linear, non-linear and time frequency features, in order to develop the machine learning model. They found that the combined machine learning approach performed better than traditional machine learning classifiers (Multilayer Perceptron, K-Nearest Neighbor, Support Vector Machine) (accuracy of 98.21% vs. 91.90–93.76%, respectively) (Ebrahimzadeh et al., 2018).

Thus, there are numerous techniques in machine learning being explored for the use of electrophysiological data to predict AF. These techniques range from different "traditional" machine learning classification algorithms to convolutional neural networks. No direct comparison between traditional models and machine learning models in ECG interpretation have been performed to date. However, machine learning methodologies allow analysis of large quantities of ECG data that may be too cumbersome and time consuming to perform manually and has thus far allowed for the development of prediction models with strong diagnostic performance.

FUTURE DIRECTION

As the reader considers the various sections in this review from clinical data to electrophysiological data, we can see that machine learning, while still in its infancy, has begun to drastically improve our ability to predict AF. There are current worldwide efforts and clinical trials to prospectively test and harness the potential of AI in clinical practice for AF. In the United States, the Batch Enrollment for AI-Guided Intervention to Lower Neurologic Events in Unrecognized AF, or BEAGLE trial, seeks to assess the performance of AI on detecting AF on normal sinus ECGs in adult patients who do not have a previous diagnosis of AF and are eligible for anticoagulation based on standard stroke risk stratification (Yao et al., 2021). Similar efforts are being undertaken in France, United Kingdom, the Netherlands, Finland, and Germany, some also testing the utility of AI applied to ECGs obtained by portable devices (ClinicalTrials.gov, 2021a,b,c,d,e).

Despite these important advancements, there is still significant room for growth within this space.

(1) Integration of all modalities of data: While siloed approaches are often necessary in the beginning to refine specific methodologies as it pertains to different modalities of data, we have seen from conventional studies that the combination of data (e.g., clinical, laboratory, imaging, etc.) often leads to the highest predictive ability for any clinical risk score. This same principle should be applied to machine learning algorithms, such that the development of a machine learning algorithm that can incorporate all modalities of data, likely further enhancing the powerful predictive performance of the existing AI algorithms.

(2) Advancements in our understanding and methodologies of machine learning: At this early stage, due to the nature of many advanced types of machine learning, including convolutional neural networks, the signal features selected by the AI as important predictive features in an algorithm cannot be known (the so-called "black box"). It is possible that future techniques will allow the algorithms to be more explicit and informative about its own methodologies, both to inform clinicians on novel patterns that may advance human understanding but also to inform researchers on potential troubleshooting issues, such as the inadvertent use of non-medical or unrelated data in their predictive algorithms.

(3) Implications of machine learning algorithms on management: While the overall aim of this review is to evaluate the role of AI in predicting AF, future studies should undoubtedly evaluate the prospective use of these algorithms to determine optimal management strategies for patients. In AF, for example, there is significant implication with AF diagnosis regarding stroke prevention via the use of anticoagulation. Could there be important changes to clinical outcomes and patient management based on the results of the algorithm that can be eventually be actionable, perhaps even before a clinical diagnosis?

CONCLUSION

There is no doubt that artificial intelligence will play a greater role in medicine as the technology continues to advance and our understanding of its applications continues to grow. While still in its early stages and still flawed by inherent limitations, machine learning shows great promise in improving our ability to predict AF. The future integration of clinical, imaging and electrophysiological data will certainly improve the performance of these machine learning algorithms, and ultimately improve the care of patients worldwide.

AUTHOR CONTRIBUTIONS

PN formulated the concept and format of the article and made critical revisions to the manuscript. AT wrote and revised the manuscript. Both authors contributed to the article and approved the submitted version.

REFERENCES

Alonso, A., Krijthe, B. P., Aspelund, T., Stepas, K. A., Pencina, M. J., Moser, C. B., et al. (2013). Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J. Am. Heart Assoc.* 2:e000102. doi: 10.1161/JAHA.112.000102

Atta-Fosu, T., LaBarbera, M., Ghose, S., Schoenhagen, P., Saliba, W., Tchou, P. J., et al. (2021). A new machine learning approach for predicting likelihood of recurrence following ablation for atrial fibrillation from CT. *BMC Med. Imaging.* 21:45. doi: 10.1186/s12880-021-00578-4

Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., et al. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 394, 861–867. doi: 10.1016/S0140-6736(19)31721-0

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn.* 30, 1145–1159. doi: 10.1016/S0031-3203(96)00142-2

Caputo, M., and Mondillo, S. (2012). Echocardiography in the prediction of atrial fibrillation recurrence: a review. *J. Atr. Fibrillation.* 5:675.

Chamberlain, A. M., Agarwal, S. K., Folsom, A. R., Soliman, E. Z., Chambless, L. E., Crow, R., et al. (2011). A clinical risk score for atrial fibrillation in a biracial prospective cohort (from the Atherosclerosis Risk in Communities [ARIC] study). *Am. J. Cardiol.* 107, 85–91. doi: 10.1016/j.amjcard.2010.08.049

Chang, A., Cadaret, L. M., and Liu, K. (2020). Machine learning in electrocardiography and echocardiography: technological advances in clinical cardiology. *Curr. Cardiol. Rep.* 22:161. doi: 10.1007/s11886-020-01416-9

ClinicalTrials.gov (2021a). *Artificial Intelligence Use for the Detection of Atrial Fibrillation Drivers.* Available online at: https://ClinicalTrials.gov/show/NCT03401931 (accessed June 1, 2021).

ClinicalTrials.gov (2021c). *Predicting Patient-level New Onset Atrial Fibrillation.* Available online at: https://ClinicalTrials.gov/show/NCT04657900 (accessed June 1, 2021).

ClinicalTrials.gov (2021e). *Portable Measurement Methods Combined With Artificial Intelligence in Detection of Atrial Fibrillation.* Available online at: https://ClinicalTrials.gov/show/NCT04917653 (accessed June 1, 2021).

ClinicalTrials.gov (2021b). *Exploratory Study of Apple Watch Electrocardiographic Diagnostic Performance Enhanced by an Artificial Intelligence Algorithm.* Available online at: https://ClinicalTrials.gov/show/NCT04792905 (accessed June 1, 2021).

ClinicalTrials.gov (2021d). *Prevention of Stroke and Sudden Cardiac Death by Recording of 1-Channel Electrocardiograms.* Available online at: https://ClinicalTrials.gov/show/NCT04637230 (accessed June 1, 2021).

De Vos, C. B., Weijs, B., Crijns, H. J., Cheriex, E. C., Palmans, A., Habets, J., et al. (2009). Atrial tissue Doppler imaging for prediction of new-onset atrial fibrillation. *Heart* 95, 835–840. doi: 10.1136/hrt.2008.148528

Dilaveris, P. E., Gialafos, E. J., Andrikopoulos, G. K., Richter, D. J., Papanikolaou, V., Poralis, K., et al. (2000). Clinical and electrocardiographic predictors of recurrent atrial fibrillation. *Pacing Clin. Electrophysiol.* 23, 352–358. doi: 10.1111/j.1540-8159.2000.tb06761.x

Ebersberger, U., Bernard, M. L., Schoepf, U. J., Wince, W. B., Litwin, S. E., Wang, Y., et al. (2020). Cardiac computed tomography for atrial fibrillation patients undergoing ablation: implications for the prediction of early recurrence. *J. Thorac. Imaging* 35, 186–192. doi: 10.1097/RTI.0000000000000425

Ebrahimzadeh, E., Kalantari, M., Joulani, M., Shahraki, R. S., Fayaz, F., and Ahmadi, F. (2018). Prediction of paroxysmal Atrial Fibrillation: a machine learning based approach using combined feature vector and mixture of expert classification on HRV signal. *Comput. Methods Programs Biomed.* 165, 53–67. doi: 10.1016/j.cmpb.2018.07.014

Firouznia, M., Feeny, A. K., LaBarbera, M. A., McHale, M., Cantlay, C., Kalfas, N., et al. (2021). Machine learning-derived fractal features of shape and texture of the left atrium and pulmonary veins from cardiac computed tomography scans are associated with risk of recurrence of atrial fibrillation postablation. *Circ. Arrhythm. Electrophysiol.* 14:e009265. doi: 10.1161/CIRCEP.120.009265

FitzHenry, F., Resnic, F. S., Robbins, S. L., Denton, J., Nookala, L., Meeker, D., et al. (2015). Creating a common data model for comparative effectiveness with the observational medical outcomes partnership. *Appl. Clin. Inform.* 6, 536–547. doi: 10.4338/ACI-2014-12-CR-0121

Fu, G. S., Levin-Schwartz, Y., Lin, Q. H., and Zhang, D. (2019). Machine learning for medical imaging. *J. Healthcare Eng.* 2019:9874591. doi: 10.1155/2019/9874591

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.* Newton, MA: O'Reilly Media.

Hill, N. R., Ayoubkhani, D., McEwan, P., Sugrue, D. M., Farooqui, U., Lister, S., et al. (2019). Predicting atrial fibrillation in primary care using machine learning. *PLoS One* 14:e0224582. doi: 10.1371/journal.pone.0224582

Himmelreich, J. C. L., Lucassen, W. A. M., Harskamp, R. E., Aussems, C., van Weert, H., and Nielen, M. M. J. (2021). CHARGE-AF in a national routine primary care electronic health records database in the Netherlands: validation for 5-year risk of atrial fibrillation and implications for patient selection in atrial fibrillation screening. *Open Heart* 8:e001459. doi: 10.1136/openhrt-2020-001459

Hirose, T., Kawasaki, M., Tanaka, R., Ono, K., Watanabe, T., Iwama, M., et al. (2012). Left atrial function assessed by speckle tracking echocardiography as a predictor of new-onset non-valvular atrial fibrillation: results from a prospective study in 580 adults. *Eur. Heart J. Cardiovasc. Imaging* 13, 243–250. doi: 10.1093/ejechocard/jer251

Hu, W. S., and Lin, C. L. (2020). Prediction of new-onset atrial fibrillation for general population in Asia: a comparison of C2HEST and HATCH scores. *Int. J. Cardiol.* 313, 60–63. doi: 10.1016/j.ijcard.2020.03.036

Hwang, S. H., Oh, Y. W., Lee, D. I., Shim, J., Park, S. W., and Kim, Y. H. (2015). Relation between left atrial wall composition by late gadolinium enhancement and complex fractionated atrial electrograms in patients with persistent atrial fibrillation: influence of non-fibrotic substrate in the left atrium. *Int. J. Cardiovasc. Imaging* 31, 1191–1199. doi: 10.1007/s10554-015-0675-1

Im, S. I., Park, D. H., Kim, B. J., Cho, K. I., Kim, H. S., and Heo, J. H. (2018). Clinical and electrocardiographic characteristics for prediction of new-onset atrial fibrillation in asymptomatic patients with atrial premature complexes. *Int. J. Cardiol. Heart Vasc.* 19, 70–74. doi: 10.1016/j.ijcha.2018.05.002

Koohy, H. (2017). The rise and fall of machine learning methods in biomedical research. *F1000Res.* 6:2012. doi: 10.12688/f1000research.13016.1

Li, Y. G., Pastori, D., Farcomeni, A., Yang, P. S., Jang, E., Joung, B., et al. (2019). A Simple Clinical Risk Score (C2HEST) for Predicting Incident Atrial Fibrillation in Asian Subjects: Derivation in 471,446 Chinese Subjects. With Internal Validation and External Application in 451,199 Korean Subjects. *Chest* 155, 510–518. doi: 10.1016/j.chest.2018.09.011

Lip, G. Y. H., Skjoth, F., Nielsen, P. B., and Larsen, T. B. (2020). Evaluation of the C2HEST Risk Score as a Possible Opportunistic Screening Tool for Incident Atrial Fibrillation in a Healthy Population (From a Nationwide Danish Cohort Study). *Am. J. Cardiol.* 125, 48–54. doi: 10.1016/j.amjcard.2019.09.034

Nakamura, K., Funabashi, N., Uehara, M., Ueda, M., Murayama, T., Takaoka, H., et al. (2011). Left atrial wall thickness in paroxysmal atrial fibrillation by multislice-CT is initial marker of structural remodeling and predictor of transition from paroxysmal to chronic form. *Int. J. Cardiol.* 148, 139–147. doi: 10.1016/j.ijcard.2009.10.032

Nakatani, Y., Sakamoto, T., Yamaguchi, Y., Tsujino, Y., Kataoka, N., and Kinugawa, K. (2020). Left atrial wall thickness is associated with the low-voltage area in patients with paroxysmal atrial fibrillation. *J. Interv. Card. Electrophysiol.* 58, 315–321. doi: 10.1007/s10840-019-00611-1

O'Neal, W. T., Venkatesh, S., Broughton, S. T., Griffin, W. F., and Soliman, E. Z. (2016). Biomarkers and the prediction of atrial fibrillation: state of the art. *Vasc. Health Risk Manag.* 12, 297–303. doi: 10.2147/VHRM.S75537

Schnabel, R. B., Sullivan, L. M., Levy, D., Pencina, M. J., Massaro, J. M., D'Agostino, R. B. Sr., et al. (2009). Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *Lancet* 373, 739–745. doi: 10.1016/S0140-6736(09)60443-8

Sekelj, S., Sandler, B., Johnston, E., Pollock, K. G., Hill, N. R., Gordon, J., et al. (2021). Detecting undiagnosed atrial fibrillation in UK primary care: validation of a machine learning prediction algorithm in a retrospective cohort study. *Eur. J. Prev. Cardiol.* 28, 598–605. doi: 10.1177/2047487320942338

Shulman, E., Kargoli, F., Aagaard, P., Hoch, E., Di Biase, L., Fisher, J., et al. (2016). Validation of the Framingham heart study and CHARGE-AF risk scores for

atrial fibrillation in Hispanics, African-Americans, and Non-Hispanic Whites. *Am. J. Cardiol.* 117, 76–83. doi: 10.1016/j.amjcard.2015.10.009

Siebermair, J., Suksaranjit, P., McGann, C. J., Peterson, K. A., Kheirkhahan, M., Baher, A. A., et al. (2019). Atrial fibrosis in non-atrial fibrillation individuals and prediction of atrial fibrillation by use of late gadolinium enhancement magnetic resonance imaging. *J. Cardiovasc. Electrophysiol.* 30, 550–556. doi: 10.1111/jce.13846

Sinner, M. F., Stepas, K. A., Moser, C. B., Krijthe, B. P., Aspelund, T., Sotoodehnia, N., et al. (2014). B-type natriuretic peptide and C-reactive protein in the prediction of atrial fibrillation risk: the CHARGE-AF Consortium of community-based cohort studies. *Europace* 16, 1426–1433. doi: 10.1093/europace/euu175

Siontis, K. C., Yao, X., Pirruccello, J. P., Philippakis, A. A., and Noseworthy, P. A. (2020). How will machine learning inform the clinical care of atrial fibrillation? *Circ. Res.* 127, 155–169. doi: 10.1161/CIRCRESAHA.120.316401

Stankovic, L., and Mandic, D. (2021). Convolutional neural networks demystified: a matched filtering perspective based tutorial. *arXiv* [Preprint]. Available online at: https://arxiv.org/pdf/2108.11663.pdf (accessed June 1, 2021).

Suenari, K., Chao, T. F., Liu, C. J., Kihara, Y., Chen, T. J., and Chen, S. A. (2017). Usefulness of HATCH score in the prediction of new-onset atrial fibrillation for Asians. *Medicine (Baltimore).* 96, e5597. doi: 10.1097/MD.0000000000005597

Sultana, N., Kamatham, Y., and Kinnara, B. (eds) (2016). "Performance analysis of artificial neural networks for cardiac arrhythmia detection," in *Proceedings of the 2016 IEEE 6th International Conference on Advanced Computing (IACC)*, Bhimavaram. doi: 10.1109/IACC.2016.84

Szilagyi, J., Walters, T. E., Marcus, G. M., Vedantham, V., Moss, J. D., Badhwar, N., et al. (2018). Surface ECG and intracardiac spectral measures predict atrial fibrillation recurrence after catheter ablation. *J. Cardiovasc. Electrophysiol.* 29, 1371–1378. doi: 10.1111/jce.13699

Thong, T., McNames, J., Aboy, M., and Goldstein, B. (2004). Prediction of paroxysmal atrial fibrillation by analysis of atrial premature complexes. *IEEE Trans. Biomed. Eng.* 51, 561–569. doi: 10.1109/TBME.2003.821030

Thrun, S., and Littman, M. L. (2000). *Reinforcement Learning: An Introduction.* Cambridge, MA: The MIT Press.

Tiwari, P., Colborn, K. L., Smith, D. E., Xing, F., Ghosh, D., and Rosenberg, M. A. (2020). Assessment of a machine learning model applied to harmonized electronic health record data for the prediction of incident atrial fibrillation. *JAMA Netw. Open* 3:e1919396. doi: 10.1001/jamanetworkopen.2019.19396

Uddin, S., Khan, A., Hossain, M. E., and Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction.

BMC Med. Inform. Decis. Mak. 19:1–16. doi: 10.1186/s12911-019-1004-8

Xu, H. F., He, Y. M., Qian, Y. X., Zhao, X., Li, X., and Yang, X. J. (2011). Left ventricular posterior wall thickness is an independent risk factor for paroxysmal atrial fibrillation. *West. Indian Med. J.* 60, 647–652.

Yao, X., Attia, Z. I., Behnken, E. M., Walvatne, K., Giblon, R. E., Liu, S., et al. (2021). Batch enrollment for an artificial intelligence-guided intervention to lower neurologic events in patients with undiagnosed atrial fibrillation: rationale and design of a digital clinical trial. *Am. Heart J.* 239, 73–79. doi: 10.1016/j.ahj.2021.05.006

Yoshizawa, T., Niwano, S., Niwano, H., Igarashi, T., Fujiishi, T., Ishizue, N., et al. (2014). Prediction of new onset atrial fibrillation through P wave analysis in 12 lead ECG. *Int. Heart J.* 55, 422–427. doi: 10.1536/ihj.14-052