



Artificial intelligence for the detection, prediction, and management of atrial fibrillation

Jonas L. Isaksen¹ · Mathias Baumert² · Astrid N. L. Hermans³ · Molly Maleckar⁴ · Dominik Linz^{1,3}

¹ Department of Biomedical Sciences, University of Copenhagen, Copenhagen, Denmark

² School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide, Australia

³ Department of Cardiology, Maastricht University Medical Center and Cardiovascular Research Institute Maastricht, Maastricht, The Netherlands

⁴ Department of Computational Physiology, Simula Research Laboratory, Oslo, Norway

In this article

- Background
- Introduction to artificial intelligence
- Evaluation of the performance of AI algorithms
- AI development
- AF detection with AI
- AF prediction with AI
- AF management with AI
- Inherent bias and implications for implementation
- Conclusions

Abstract

The present article reviews the state of the art of machine learning algorithms for the detection, prediction, and management of atrial fibrillation (AF), as well as of the development and evaluation of artificial intelligence (AI) in cardiology and beyond. Today, AI detects AF with a high accuracy using 12-lead or single-lead electrocardiograms or photoplethysmography. The prediction of paroxysmal or future AF currently operates at a level of precision that is too low for clinical use. Further studies are needed to determine whether patient selection for interventions may be possible with machine learning.

Keywords

AF · AI · Machine learning · Neural networks · Disease management · Deep learning

Background

Atrial fibrillation (AF) is the most common sustained arrhythmia and affects more than 43 million people globally [1]. In the European Union, almost 8 million people >65 years of age had AF in 2016, a number that is expected to increase to over 14 million by 2060 due to increased longevity and increasing prevalence of AF risk factors, which leads to increased costs associated with detection, diagnosis, and management of AF. Within the first year of diagnosis, each AF patient in Germany is associated with a cost of over 2200 € [2]. Screening efforts are costly: data from the Gutenberg Health Study estimated 12-lead electrocardiogram (ECG)-based screening in the 65- to 74-year-old general population to cost approximately 30,000 € per gained quality-adjusted life-year [3].

Artificial intelligence (AI) methods—including machine learning and artificial neural networks (deep learning)—can per-

form some tasks much faster than human experts and at a level comparable to humans or with even greater precision [4–6]. Nevertheless, recent experience suggests that the performance of AI-based systems often fails once they are implemented in a real-life setting, underlining the importance of a careful approach to AI development and validation [7, 8]. Although AI methods have the potential to improve AF detection and even to reduce costs [9], the exact role of AI in clinical AF management remains unclear.

Thus, the aim of this review article was to summarize the state of the art in AI-based detection and management of AF and to review the important steps necessary for development of AI-based systems.

Introduction to artificial intelligence

The term “artificial intelligence” was coined at the 1956 summer research workshop at Dartmouth College (Hanover, NH, USA)



Scan QR code & read article online

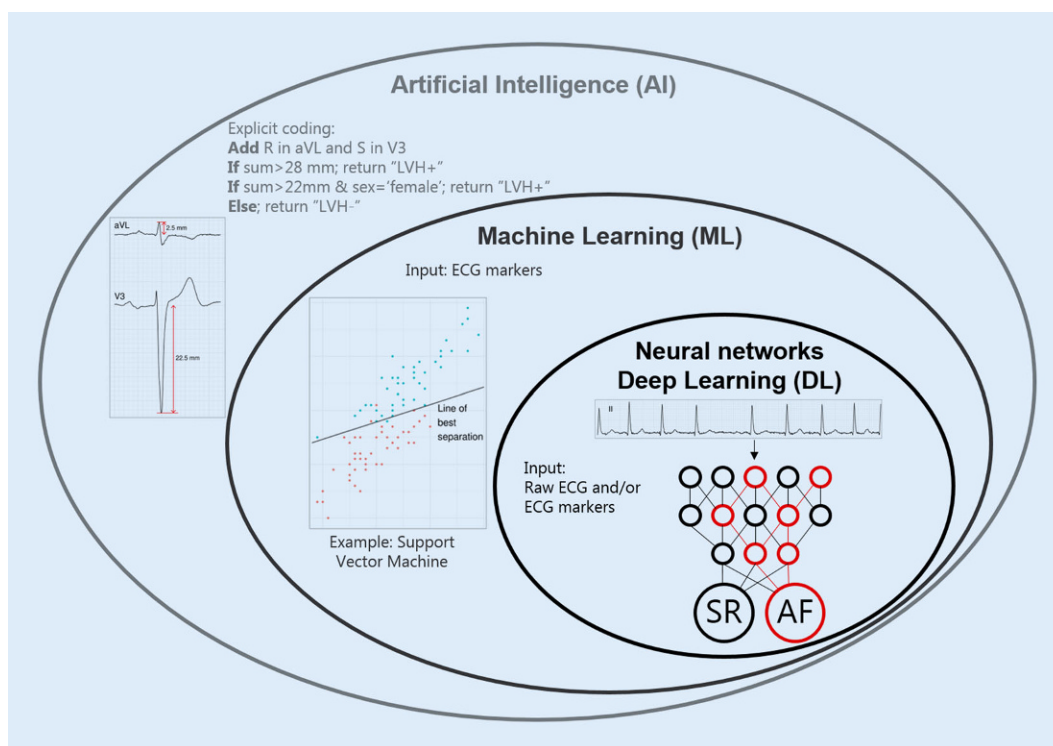


Fig. 1 ◀ Distinction and overlap between artificial intelligence (AI), machine learning (ML), and neural networks/deep learning (DL). Only DL operates directly on raw waveforms, while other AI methods rely on markers or features. *AI example:* explicit programming to determine whether left ventricular hypertrophy (LVH) is present or not based on R and S amplitudes. *ML example:* (imperfect) discrimination between two classes using a support vector machine. *DL example:* Discrimination between sinus rhythm (SR) and atrial fibrillation (AF) using a simple neural network operating directly on a raw electrocardiogram (ECG) rhythm strip

[10], 6 years after Turing had asked the question “Can machines think?” [11]. The 1956 workshop aimed to develop a machine that could pass the Turing test, i.e., exhibit intelligence at the level of humans. Machines that actually think have not yet been developed, but AI also includes machines that perform simulated thinking, i.e., solve tasks that would otherwise require human intelligence. As such, an automated ECG interpretation software coded line-by-line is within the field of AI (■ Fig. 1). Machine learning is the subset of AI that relates to self-learning algorithms trained with data, although at times AI is used synonymously with machine learning. Deep learning is a special case of machine learning, whereby information is passed from layer to layer of artificial neurons in a neural network, a process that was inspired by human neuronal computing [12]. The main advantages of artificial neural networks over other machine learning methods are that extraction of specific features is not needed and that the networks are flexible to comprehend complex data. The main limitation of neural networks over other machine learning methods is the need for large, annotated data sets for the model to be successful (i.e., to converge). Many different neural

architectures exist; of special interest to AF are recurrent neural networks (RNNs) and convolutional neural networks (CNNs). RNNs are ideal for long time series and forecasting, whereas CNNs are used for image and signal analysis [6, 7, 12]. Many papers referenced in this review used CNNs.

Although AI technically includes programs that are not classified as machine learning, this review will focus on the machine learning part of AI algorithms, and special emphasis is put on neural networks (deep learning).

Evaluation of the performance of AI algorithms

Metrics for evaluation of machine learning algorithms are the same as for any other algorithm or diagnostic test, however often a condition may be rare, and researchers should be aware of pitfalls among some metrics in this case.

For classification tasks, the standard confusion matrix-derived metrics of sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are useful. Accuracy (the share of correct predictions out of total predictions) is a good metric only in balanced datasets, but may not reflect the diagnostic perfor-

mance in imbalanced datasets (i.e., rare events). To illustrate this issue: if an event occurs in one out of 1000 people, a model that predicts the event never to occur would achieve a seemingly good accuracy of 99.9%. In imbalanced datasets, model performance may be suboptimal despite high sensitivity and specificity. One paper reported a sensitivity of 79.0% and a specificity of 79.5% for the prediction of AF, but the PPV (not reported) was 26.1% [13]. For imbalanced datasets, the PPV should be reported along with sensitivity and specificity, and the accuracy metric should be replaced by the F1 score, which is the harmonic mean of sensitivity and PPV and may more accurately reflect the performance of the model [4, 14]. In the aforementioned paper, the reported accuracy and the F1 score were 79.4 and 39.2%, respectively. The area under that receive-operator characteristics (ROC) curve (AUC), also termed the C-statistics, is based on sensitivity and specificity and is unsuitable for algorithms evaluated on imbalanced datasets [14] (in the previously mentioned paper [13], AUC was 0.87). The ROC curve should be replaced by or supplemented with the precision-recall curve (PRC), whereby the axes are precision (PPV) and recall (sensitivity) (see ■ Fig. 2). The ROC-based AUC/C-

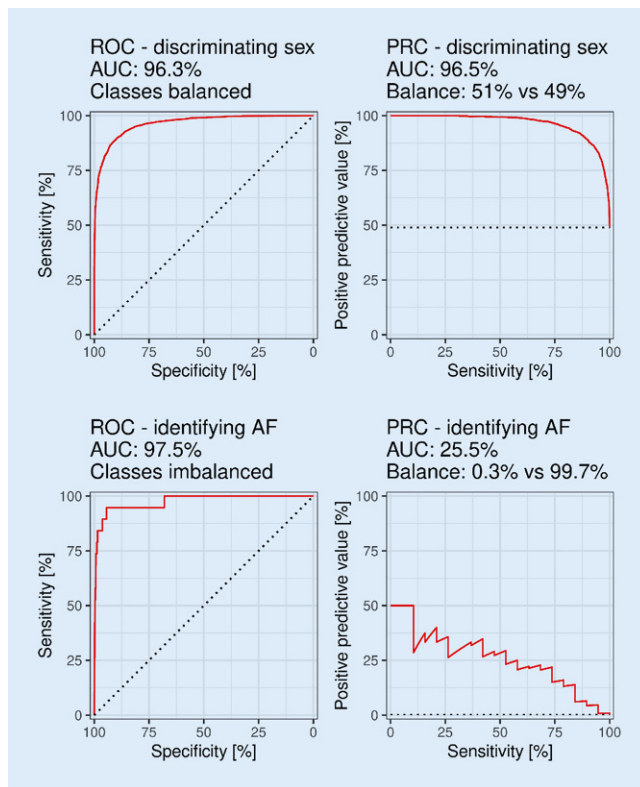


Fig. 2 ◀ *Top*: Model performance is similarly displayed by the receiver-operator characteristics (ROC) curve and the precision-recall curve (PRC) when classes are balanced. *Bottom*: Poor model performance combined with imbalanced classes results in a good-looking ROC, but the PRC curve reveals the poor performance. For PRC curves, positive predictive value is often termed precision, and sensitivity termed recall. (AF Atrial fibrillation, AUC area under the curve)

statistic should be replaced by the corresponding area under the PRC (sometimes termed AUPRC) [14].

For regression tasks, the prediction error is often quantified using the mean absolute error (MAE) or the root-mean-squared error (RMSE) [6]. The RMSE metric punishes outliers more heavily than the MAE metric, and a large difference between MAE and RMSE thus indicates that extreme outliers exist, which may be caused by catastrophic predictions or large ground truth errors. The prediction error may also be visualized with a “predicted vs. ground truth” scatter plot including the identity line, often accompanied by a correlation coefficient, or a Bland–Altman plot [15].

AI development

To develop machine learning models, a large dataset is typically required. In supervised machine learning (i.e., with labelled data), a dataset is typically split into two parts: training and testing. This is a crucial step to avoid over-fitting, which is the phenomenon whereby the model is not just trained to use the overall characteristics of a biological sig-

nal or measurement, but starts to infer relations to the underlying population. A consequence of over-fitting is that the model will perform poorly in other unseen populations.

Another frequently used strategy to avoid over-fitting is cross-validation, whereby the data is split randomly into a number of subsets, for instance, five equally sized sets. The same underlying model is then trained in five replications, each using a different 20% of the data for testing/validation and the remaining 80% of the data for training. This method uses the entire dataset, is unbiased, and furthermore allows the researcher to assess model stability.

Despite the use of train-test split or cross-validation, over-fitting may occur if the training is done repeatedly and the test error is used to select the optimal hyperparameters (fine tuning for machine learning) or to optimize network architecture (deep learning). In this case, a validation set should be identified to assess the performance of the final model and to ensure the collection of the correct metrics.

Regardless of the training strategy, value is added if machine learning models are validated in an external population

to show that they generalize well. Even though this step does not guarantee that AI works as intended in general practice, it arguably improves the likelihood that the developed model will be useful upon implementation [7].

AF detection with AI

The current guidelines of the European Society of Cardiology on the Diagnosis and Management of Atrial Fibrillation require a 12-lead ECG or >30-s single-lead ECG documentation for a definitive diagnosis of AF [1]. However, the guidelines also highlight the need for screening, risk tools, and prediction models for AF and specifically mention the use of both mobile health (mHealth) technology and AI.

AF detection using AI works fairly well on 12-lead ECGs, and the main source of false-positive detections is premature atrial contractions and pronounced respiratory arrhythmia, which frequently occurs in young individuals. Using standard 12-lead ECGs, Cai et al. applied neural networks to obtain an F1 score of 95–96% (99% for AF vs sinus rhythm, excluding other rhythms) [16]. Jo et al. developed a neural network with a sensitivity of 98.5% and a PPV of 95.4% [17]. They furthermore showed good generalization to external datasets and achieved sensitivities of 99.6–99.9% and PPVs of 91.4–98.0%. Interestingly, the performance of the model was only moderately reduced using lead I only, suggesting that this model may also be used for wearables. Baalman et al. used extensive pre-processing to single heart cycles and then found that lead II and lead V3 were the best candidates for single-lead AF detection using deep learning [18]. However, this strategy only achieved F1 scores of at most 94%. Despite good performance, a definite diagnosis of AF currently needs to be confirmed by a physician. However, 12-lead AI may serve as an initial filter to reduce the number of false-positives, as was done with success for implantable loop recorders [19].

Single-lead ECG recordings for AF detection (see ◼ Fig. 3) are gaining more interest, since they can be performed by several available mHealth devices and wearables [20]. Hannun et al. obtained 91,232 30-s, single-lead ECGs from patch devices

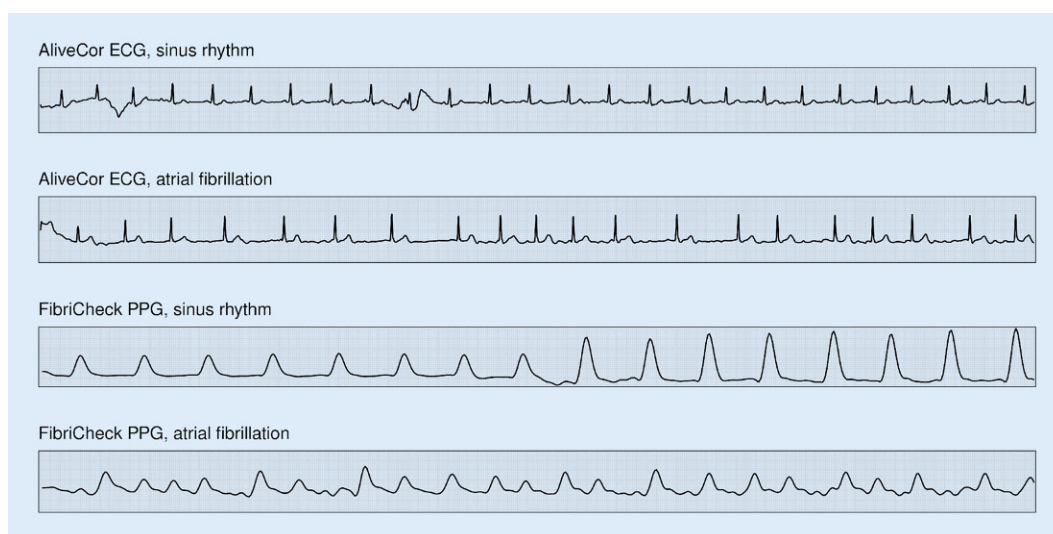


Fig. 3 ◀ Examples of sinus rhythm and atrial fibrillation recorded using a single lead KardiaMobile (AliveCor) electrocardiogram (ECG) device (top) and a smartphone-based FibriCheck photoplethysmography (PPG, bottom). Recordings are from different individuals

on >50,000 people and trained a neural network to identify 12 unique rhythms [4]. In a test set of 328 patients, they showed that their network performed better than a group of cardiologists, for multiple arrhythmias including AF. For the 2017 Computing in Cardiology (CinC) Challenge, teams competed to provide the best AF detection among 3658 hidden recordings or varying lengths made with a KardiaMobile single-lead device (a training set with 8528 similar ECGs were available) [21]. There are four classes; sinus rhythm, AF, other rhythm, and noise. Most algorithms generalized with a drop in F1 score (indicating over-fitting to the training set) to a maximum F1 of 83.1%. A voting approach among competing algorithms was better than any individual algorithm and improved F1 to 86.7%. In 2021, a combination of state-of-the-art algorithms improved the F1 score to 90% [22]. The CinC Challenge dataset is freely available and has been frequently used for algorithm development or validation [23–25].

Single-lead ECG may be integrated into a smartwatch wristband, making it widely available, and one study found a high sensitivity (97.7%) albeit a low PPV (40%) validated against an implantable cardiac monitor [26]. Cloud-based AI analysis of wearable ECG may allow continuous updates to the algorithm and facilitate contact with healthcare personnel [27]. Deep learning appears to be a stronger tool than classical machine learning for single-lead ECG analysis including detection of AF [28], and

single-lead ECG may play a prominent role in AF detection in the future.

Photoplethysmography (PPG) allows continuous measurement of heart rate using a smartwatch or a smartphone camera through the delineation of the pulse wave (see ◼ Fig. 3) [29]. Chen et al. compared wristband PPG and ECG measurements and found slightly higher sensitivity and PPV for the ECG-based measurements [30]. Although their neural network was faster than cardiologists at analyzing wristband PPG or ECG, it was not significantly better. Tison et al. evaluated a smartwatch-based PPG against a 12-lead ECG in 51 patients undergoing cardioversion with a sensitivity of 98% and a PPV of 91% [31]; interestingly, the authors showed that feeding the raw sensor data to a neural network was better than feeding (processed) heart rate data only [32]. Poh et al. similarly evaluated a neural network on 1013 patients with a sensitivity of 95.2% and a PPV of 72.3%, which increased to 100% sensitivity and 87.5% PPV with three measurements from each patient [33]. Taking a clever approach to neural network analysis, whereby they first trained a network to compress the wristband PPG signals and then repurposed the network for AF detection, Torres-Soto and Ashley got an F1 score of 96% [34]. Taken together, AI-powered PPG-based detection of AF is possible today with an acceptable false-positive rate and should now be tested in clinical and daily practice [35].

AI was also used for AF detection in other modalities. Two studies have demonstrated that AI may enable AF detection using video [36, 37]. By extracting a facial PPG signal, they achieved sensitivities of 94% and PPVs of 90–98% in a controlled setting. In a small study of 59 patients with paroxysmal AF, Jiang et al. used AI and ballistocardiograms (vibration measurements on the torso) to detect AF with a sensitivity of 96% and a PPV of 94%, although without external validation [38].

Detection of AF will likely improve in the future with the addition of novel algorithms and additional data sets for development—and with clinical validation of algorithms, we may soon see use in clinical practice. The PRICE study plans to enroll 100,000 patients to develop and validate multi-class arrhythmia detection including AF, and the investigators plan a prospective validation of the algorithm [39]. The Scripps Clinic enrolled 25,458 patients with 12-lead ECGs to train a network for multi-class arrhythmia detection [40]. In a large study with an expected 168,000 participants, patients are randomized to access or no access to AI-ECG interpretation [41]. This study will help quantify the added value of AI in clinical practice in a low-risk setting. Three studies focus on AF detection among people with increased stroke risk. One study is enrolling 2450 patients to transfer an existing AI model to work with the S-Patch Cardio and then uses the device to detect AF in patients with high stroke risk [42]. The HUA-TUO AF Trial will

randomize 1740 stroke survivors without documented AF to a handheld single-lead ECG analyzed by an AI algorithm or to usual care [43]. This study will assess the combination of a device and AI interpretation on AF detection and stroke recurrence in a high-risk population. A Leicester, United Kingdom, group is evaluating whether AI in combination with simulation may help discriminate strokes caused by AF, using magnetic resonance imaging during admission for stroke [44]. However, that study may be underpowered since it plans to enroll only 100 patients.

AF prediction with AI

One step further than AF detection is AF prediction. Based on clinical risk factors alone, machine learning was unable to improve risk prediction beyond the performance of the CHARGE-AF risk score [45]. Also for post-operative AF, machine learning was not better than a logistic regression model when applied to clinical variables [46]. The FIND-AF study set out to detect new-onset AF using AI and electronic health records in a database of 140,000 patients, and perhaps they can show that neural networks can be successfully applied to risk factors [47].

The added value of AI within prediction probably lies in neural network analysis of ECGs, as well as other signals, and the combination with clinical risk factors. Khurshid et al. used AI analysis of ECGs to predict 5-year risk of AF at a level comparable to the CHARGE-AF risk score, but the combination of the ECG and risk factors identified those at highest risk of AF [48]. Although predictions improved relatively, there is still a long way to go. With a sensitivity of 80%, the PPV in the internal validation set and the two external validation sets was approximately 17, 12, and 2%, respectively. As described previously, another neural network was able to distinguish paroxysmal AF from no AF using sinus rhythm ECGs with a sensitivity of 79% and a PPV of 26% [13]. When the authors applied that same model to patients with embolic stroke of unknown origin, after changing the detection threshold, they got a similar PPV of 23%, but the sensitivity dropped to 63% [49]. The model may thus help predict incident AF, and this is

currently being validated prospectively in the BEAGLE study [50].

AF management with AI

A number of studies have developed machine learning algorithms to assist physicians of different specialties in the management of AF. One group developed an app with a machine learning backend that improved identification of AF in the emergency department and recommended appropriate anticoagulant treatment [51]. However, Levy et al. were unable to train a machine learning algorithm to dose dofetilide as physicians did, perhaps due to the complexity of the clinical situation and perhaps due to large variations between individual physicians [52].

Efforts have been made on risk stratification. Inohara et al. used unsupervised machine learning to cluster patients into groups that proved to have slightly but significantly different risks of major adverse cardiovascular or neurological events. Wanatabe et al. used machine learning on clinical risk factors to predict thromboembolisms slightly better than a logistic regression model [53]. For outcomes of major bleeding and mortality, respectively, the model was not better than logistic regressions. Samaras et al. developed a similar machine learning algorithm for the outcome of mortality, but they did not compare it to logistic regression, nor did they validate the algorithm externally [54]. Loring et al. analyzed machine learning efforts compared to traditional regression for outcomes of death, bleeding, and stroke, and did not find an added value of machine learning [55]. The research so far does not indicate that machine learning can empower clinical variables much beyond classical point scores or logistic regression.

Signal analysis using machine learning may help identify responders to treatment since the complexity of the signals may be better handled by a neural network than human explicit signal processing. So far, in two studies in patients, machine learning did not aid in identifying responders to direct current cardioversion [56] or pulmonary vein isolation [57], respectively; however, both studies had enrolled

too few patients for algorithm development. The AI-PAFA Trial will prospectively randomize 340 AF patients to be evaluated for catheter ablation using an AI algorithm or conventional guideline-based rules [58]. A 2021 simulation study indicated that machine learning algorithms may be able to identify pulmonary vein isolation responders [59]; however, further validation studies are needed. A Swedish study is evaluating the use of smartphone-based PPG to detect spontaneous conversion to sinus rhythm for patients about to undergo direct current cardioversion, which is a step towards improved patient selection [60].

Inherent bias and implications for implementation

The implementation of AI will impact society and care. AI-based tools reflect the data that they were trained on, and a bias in the training data will generally lead to a biased decision [61]. A fault analysis of a decision tool for recurrent crime risk assessment (COMPAS) showed that false-positive predictions of recurrent crimes occurred more often in black people, whereas false-negative predictions occurred more often in white people [62]. In medicine, similar systematic discrimination was seen for a system that assigned special care to certain patients, for which black patients consequently had to be much sicker compared to white patients in order to be referred [63]. In both cases, the AI system was biased since it was trained on biased data, i.e., history showed that black patients actually did have to be sicker to be referred when physicians made the referrals. Thus, bias in training data can translate into a biased AI system. However, AI may also be a powerful tool to identify and mitigate such bias [61].

AI for signal analysis such as AF detection, AF prediction, or AF management is not prone to such systematic bias compared to AI systems operating on patient record data. However, both types of analyses may be prone to data shift. Data shift is the phenomenon whereby the input parameters change, leading to a different relation between input data and outcomes, and with a corresponding change in AI accuracy [64]. One way data may “change”

is due to a shift in the underlying patient population. Under the coronavirus disease 2019 pandemic, a change in the relation between fevers and bacterial sepsis due to a change in patient population led to the breakdown of an AI system for sepsis alerting [64]. In AI for signal analysis, this may happen if the device is changed (e.g., a software update) or replaced with another device, or if screening is employed in a different population (e.g., younger individuals than used for training). Thus, continuous quality assessment is an important part of implementation of AI methods. Transfer learning is the fine-tuning of an existing algorithm [65] and may be a way to overcome the data shift issue, just as transfer learning may be used to adapt an algorithm developed in one population to operate optimally in a different population.

Conclusions

The clinical application of AI has the potential to contribute to AF care in the digital era. Neural networks applied to 12-lead or single-lead ECG or PPG recordings show good performance in detecting AF. Additionally, AI-based prediction of AF shows more potential when applied to biomedical signals than to clinical variables, but these algorithms need to be further developed. AI may even support identification of treatment responders to rhythm control strategies. The real-life performance of the algorithms must be validated in additional populations, and future studies are warranted before they can be broadly implemented in clinical pathways and screening programs. To realize the full potential of AI, the identification of useful actionable data and the integration of AI-derived information into patient management pathways and treatment decision processes represents an important research area to facilitate the clinical implementation in the future and may help to ensure a patient-oriented focus in AF care. Clinical validation, internal and external replication consistency, and generalizability in various healthcare settings regardless of available resources should be included in future research programs.

Corresponding address

Dr. Dominik Linz

Department of Biomedical Sciences, University of Copenhagen
Copenhagen, Denmark
dominik.linz@gmx.de

Funding. Open access funding provided by Maastrecht University.

Declarations

Conflict of interest. J.L. Isaksen, M. Baumert, A.N.L. Hermans, M. Maleckar, and D. Linz declare that they have no competing interests.

For this article no studies with human participants or animals were performed by any of the authors. All studies performed were in accordance with the ethical standards indicated in each case.

Open Access. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, Boriani G, Castella M, Dan G-A, Dilaveris PE et al (2021) 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European association for cardio-thoracic surgery (EACTS): the task force for the diagnosis and management of atrial fibrillation of the European society of cardiology (ESC) developed with the special contribution of the European heart rhythm association (EHRA) of the ESC. *Eur Heart J* 42:373–498
- Reinhold T, Lindig C, Willich SN, Brüggengjürgen B (2011) The costs of atrial fibrillation in patients with cardiovascular comorbidities—a longitudinal analysis of German health insurance data. *Europace* 13:1275–1280
- Schnabel RB, Wallenhorst C, Engler D, Blankenberg S, Pfeiffer N, Spruncker NA, Buettner M, Michal M, Lackner KJ, Münzel T et al (2021) Refined atrial fibrillation screening and cost-effectiveness in the German population. *Heart*. <https://doi.org/10.1136/heartjnl-2020-318882>
- Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY (2019) Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 25:65–69
- Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, Topol EJ, Ioannidis JPA, Collins GS, Maruthappu M (2020) Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* 368:m689
- Hicks SA, Isaksen JL, Thambawita V, Ghouse J, Ahlberg G, Linneberg A, Grarup N, Strömke I, Ellervik C, Olesen MS et al (2021) Explaining deep neural networks for knowledge discovery in electrocardiogram analysis. *Sci Rep* 11:10949
- Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK (2018) Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Med* 15:e1002683
- Lin H, Li R, Liu Z, Chen J, Yang Y, Chen H, Lin Z, Lai W, Long E, Wu X et al (2019) Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. *EClinicalMedicine* 9:52–59
- Hill NR, Sandler B, Mokgokong R, Lister S, Ward T, Boyce U, Farrowqui U, Gordon J (2020) Cost-effectiveness of targeted screening for the identification of patients with atrial fibrillation: evaluation of a machine learning risk prediction algorithm. *J Med Econ* 23:386–393
- McCarthy J, Minsky ML, Rochester N, Corporation IBM, Shannon CE A proposal for the Dartmouth summer research project on artificial intelligence. <http://jmc.stanford.edu/articles/dartmouth.html>. Accessed 13 Jan 2022
- Turing AM (1950) Computing machinery and intelligence. *Mind* LIX:433–460
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
- Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, Carter RE, Yao X, Rabinstein AA, Erickson BJ 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
- Ozenne B, Subtil F, Maucort-Boulch D (2015) The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J Clin Epidemiol* 68:855–859
- Altman DG, Bland JM (1983) Measurement in medicine: the analysis of method comparison studies. *JR Stat Soc Ser A Stat Soc* 32:307–317
- Cai W, Chen Y, Guo J, Han B, Shi Y, Ji L, Wang J, Zhang G, Luo J (2020) Accurate detection of atrial fibrillation from 12-lead ECG using deep neural network. *Comput Biol Med* 116:103378
- Jo Y-Y, Cho Y, Lee SY, Kwon J-M, Kim K-H, Jeon K-H, Cho S, Park J, Oh B-H (2021) Explainable artificial intelligence to detect atrial fibrillation using electrocardiogram. *Int J Cardiol* 328:104–110
- Baalman SWE, Schroevers FE, Oakley AJ, Brouwer TF, van der Stuijt W, Bleijendaal H, Ramos LA, Lopes RR, Marquering HA, Knops RE et al (2020) A morphology based deep learning model for atrial fibrillation detection using single cycle electrocardiographic samples. *Int J Cardiol* 316:130–136
- Mittal S, Oliveros S, Li J, Barroyer T, Henry C, Gardella C (2021) AI filter improves positive predictive value of atrial fibrillation detection by an implantable loop recorder. *JACC Clin Electrophysiol* 7:965–975
- Hermans ANL, Gawalko M, Dohmen L, van der Velden RMJ, Betz K, Duncker D, Verhaert DVM, Heidebuchel H, Svennberg E, Neubeck L et al

- (2021) Mobile health solutions for atrial fibrillation detection and management: a systematic review. *Clin Res Cardiol*. <https://doi.org/10.1007/s00392-021-01941-9>
21. Clifford GD, Liu C, Moody B, Lehman L-WH, Silva I, Li Q, Johnson AE, Mark RG (2017) AF classification from a short single lead ECG recording: the physionet/computing in cardiology challenge 2017. *Comput Cardiol*. <https://doi.org/10.22489/CinC.2017.065-469>
 22. Rad AB, Galloway C, Treiman D, Xue J, Li Q, Sameni R, Albert D, Clifford GD (2021) Atrial fibrillation detection in outpatient electrocardiogram monitoring: an algorithmic crowdsourcing approach. *PLoS ONE* 16:e259916
 23. Zhang P, Ma C, Sun Y, Fan G, Song F, Feng Y, Zhang G (2021) Global hybrid multi-scale convolutional network for accurate and robust detection of atrial fibrillation using single-lead ECG recordings. *Comput Biol Med* 139:104880
 24. Zhang X, Li J, Cai Z, Zhang L, Chen Z, Liu C (2021) Over-fitting suppression training strategies for deep learning-based atrial fibrillation detection. *Med Biol Eng Comput* 59:165–173
 25. Oster J, Hopewell JC, Zibera K, Wijesurendra R, Camm CF, Casadei B, Tarasenko L (2020) Identification of patients with atrial fibrillation: a big data exploratory analysis of the UK biobank. *Physiol Meas* 41:25001
 26. Wasserlauf J, You C, Patel R, Valys A, Albert D, Passman R (2019) Smartwatch performance for the detection and quantification of atrial fibrillation. *Circ Arrhythm Electrophysiol* 12:e6834
 27. Tutuko B, Nurmaini S, Tondas AE, Rachmatullah MN, Darmawahyuni A, Esaferi R, Firdaus F, Sapitri AI (2021) AFibNet: an implementation of atrial fibrillation detection with convolutional neural network. *BMC Med Inform Decis Mak* 21:216
 28. Teplitzky BA, McRoberts M, Ghanbari H (2020) Deep learning for comprehensive ECG annotation. *Heart Rhythm* 17:881–888
 29. Betz K, van der Velden R, Gawalko M, Hermans A, Pluymaekers N, Hillmann HAK, Hendriks J, Duncker D, Linz D (2021) Interpretation der Photoplethysmographie: Schritt für Schritt. *Herzschrittmacherther Elektrophysiol* 32:406–411
 30. Chen E, Jiang J, Su R, Gao M, Zhu S, Zhou J, Huo Y (2020) A new smart wristband equipped with an artificial intelligence algorithm to detect atrial fibrillation. *Heart Rhythm* 17:847–853
 31. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, Vittinghoff E, Lee ES, Fan SM, Gladstone RA et al (2018) Passive detection of atrial fibrillation using a commercially available smartwatch. *JAMA Cardiol* 3:409–416
 32. Aschbacher K, Yilmaz D, Kerem Y, Crawford S, Benaron D, Liu J, Eaton M, Tison GH, Olgin JE, Li Y et al (2020) Atrial fibrillation detection from raw photoplethysmography waveforms: a deep learning application. *Heart Rhythm* 021:3–9
 33. Poh M-Z, Poh YC, Chan P-H, Wong C-K, Pun L, Leung WW-C, Wong Y-F, Wong MM-Y, Chu DW-S, Siu C-W (2018) Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms. *Heart* 104:1921–1928
 34. Torres-Soto J, Ashley EA (2020) Multi-task deep learning for cardiac rhythm detection in wearable devices. *NPJ Digit Med* 3:116
 35. Pluymaekers NAHA, Hermans ANL, van der Velden RMJ, Gawalko M, den Uijl DW, Buskes S, Vernooij K, Crijns HJGM, Hendriks JM, Linz D (2021) Implementation of an on-demand app-based heart rate and rhythm monitoring infrastructure for the management of atrial fibrillation through teleconsultation: TeleCheck-AF. *Europace* 23:345–352
 36. Yan BP, Lai WHS, Chan CKY, Au ACK, Freedman B, Poh YC, Poh M-Z (2020) High-throughput, contact-free detection of atrial fibrillation from video with deep learning. *JAMA Cardiol* 5:105–107
 37. Liu X, Yang X, Wang D, Wong A, Ma L, Li L (2021) VidAF: a motion-robust model for screening atrial fibrillation from facial videos. *IEEE J Biomed Health Inform*. <https://doi.org/10.1109/JBHI.2021.3124967>
 38. Jiang F, Hong C, Cheng T, Wang H, Xu B, Zhang B (2021) Attention-based multi-scale features fusion for unobtrusive atrial fibrillation detection using ballistocardiogram signal. *BioMed Eng OnLine* 20:12
 39. A-Rhythmik GmbH (2021) Prevention of stroke and sudden cardiac death by recording of 1-channel electrocardiograms. <https://clinicaltrials.gov/ct2/show/NCT04637230>. Accessed 12 Dec 2021
 40. Bhavnani S (2020) Development of a novel convolution neural network for arrhythmia classification for shockable cardiac rhythms. <https://clinicaltrials.gov/ct2/show/NCT03662802>. Accessed 13 Dec 2021
 41. Lin C (2021) Artificial intelligence-enable electrocardiograms system for detecting atrial fibrillation in hospital. <https://clinicaltrials.gov/ct2/show/NCT05127460>. Accessed 12 Dec 2021
 42. Yonsei University (2021) Continuous EKG monitoring using S-patch ex: prospective observational study (S-patch registry). <https://clinicaltrials.gov/ct2/show/NCT05119725>. Accessed 12 Dec 2021
 43. The University of Hong Kong (2020) Protocol for rationale and design of home-based solution for remote atrial fibrillation screening to Prevent Recurrence Stroke (HUA-TUO AF trial): a randomized open-label study. <https://clinicaltrials.gov/ct2/show/NCT04523649>. Accessed 12 Dec 2021
 44. University of Leicester (2021) Validation of 3D stroke models to differentiate stroke subtypes and predict source in acute ischaemic stroke. <https://clinicaltrials.gov/ct2/show/NCT05055960>. Accessed 13 Dec 2021
 45. Bundy JD, Heckbert SR, Chen LY, Lloyd-Jones DM, Greenland P (2020) Evaluation of risk prediction models of atrial fibrillation (from the multi-ethnic study of atherosclerosis [MESA]). *Am J Cardiol* 125:55–62
 46. Karri R, Kawai A, Thong YJ, Ramson DM, Perry LA, Segal R, Smith JA, Penny-Dimrie JC (2021) Machine learning outperforms existing clinical scoring tools in the prediction of postoperative atrial fibrillation during intensive care unit admission after cardiac surgery. *Heart Lung Circ* 30:1929–1937
 47. Nadarajah R, Wu J, Frangi AF, Hogg D, Cowan C, Gale C (2021) Predicting patient-level new-onset atrial fibrillation from population-based nationwide electronic health records: protocol of FIND-AF for developing a precision medicine prediction model using artificial intelligence. *BMJ Open* 11:e2887
 48. Khurshid S, Friedman S, Reeder C, Di Achille P, Diamant N, Singh P, Harrington LX, Wang X, Al-Alusi MA, Sarma G et al (2021) Electrocardiogram-based deep learning and clinical risk factors to predict atrial fibrillation. *Circulation*. <https://doi.org/10.1161/CIRCULATIONAHA.121.057480>
 49. Rabinstein AA, Yost MD, Faust L, Kashou AH, Latif OS, Graff-Radford J, Attia IZ, Yao X, Noseworthy PA, Friedman PA (2021) Artificial intelligence-enabled ECG to identify silent atrial fibrillation in embolic stroke of unknown source. *J Stroke Cerebrovasc Dis* 30:105998
 50. Yao X (2021) Batch enrollment for an artificial intelligence-guided intervention to lower neurologic events in patients with unrecognized atrial fibrillation (BEAGLE). <https://clinicaltrials.gov/ct2/show/NCT04208971>. Accessed 12 Dec 2021
 51. Schwab K, Nguyen D, Ungab G, Feld G, Maisel AS, Than M, Joyce L, Peacock WF (2021) Artificial intelligence MacHne learning for the detection and treatment of atrial fibrillation guidelines in the emergency department setting (AIM HIGHER): Assessing a machine learning clinical decision support tool to detect and treat non-valvular atrial fibrillation in the emergency department. *J Am Coll Emerg Physicians Open* 2:e12534
 52. Levy AE, Biswas M, Weber R, Tarakji K, Chung M, Noseworthy PA, Newton-Cheh C, Rosenberg MA (2019) Applications of machine learning in decision analysis for dose management for dofetilide. *PLoS ONE* 14:e227324
 53. Watanabe E, Noyama S, Kiyono K, Inoue H, Atarashi H, Okumura K, Yamashita T, Lip GYH, Kodani E, Origasa H (2021) Comparison among random forest, logistic regression, and existing clinical risk scores for predicting outcomes in patients with atrial fibrillation: a report from the J-RHYTHM registry. *Clin Cardiol* 44:1305–1315
 54. Samaras A, Kartas A, Akvros E, Fotos G, Dividis G, Vasdeki D, Vrana E, Rampidis G, Karvounis H, Giannakoulas G et al (2021) A novel prognostic tool to predict mortality in patients with atrial fibrillation: the BASIC-AF risk score. *Hellenic J Cardiol* 62:339–348
 55. Loring Z, Mehrotra S, Piccini JP, Camm J, Carlson D, Fonarow GC, Fox KAA, Peterson ED, Pieper K, Kakkar AK (2020) Machine learning does not improve upon traditional regression in predicting outcomes in atrial fibrillation: an analysis of the ORBIT-AF and GARFIELD-AF registries. *Europace* 22:1635–1644
 56. Vinter N, Frederiksen AS, Albertsen AE, Lip GYH, Fenger-Grøn M, Trinquart L, Frost L, Møller DS (2020) Role for machine learning in sex-specific prediction of successful electrical cardioversion in atrial fibrillation? *Open Heart* 7:e1297
 57. Shade JK, Ali RL, Basile D, Popescu D, Akhtar T, Marine JE, Spragg DD, Calkins H, Trayanova NA (2020) Preprocedure application of machine learning and mechanistic simulations predicts likelihood of paroxysmal atrial fibrillation recurrence following pulmonary vein isolation. *Circ Arrhythm Electrophysiol* 13:e8213
 58. Yonsei University (2021) Artificial intelligence guided patient selection for atrial fibrillation catheter ablation: randomized clinical trial (AI-PFA trial). <https://clinicaltrials.gov/ct2/show/NCT04997824>. Accessed 12 Dec 2021
 59. Luongo G, Azzolin L, Schuler S, Rivolta MW, Almeida TP, Martínez JP, Soriano DC, Luik A, Müller-Edenborn B, Jadidi A et al (2021) Machine learning enables noninvasive prediction of atrial fibrillation driver location and acute pulmonary vein ablation success using the 12-lead ECG. *Cardiovasc Digit Health J* 2:126–136
 60. Engdahl J (2021) Validation of a novel smartphone-based photoplethysmographic method for ambulatory heart rhythm monitoring in connection to treatment of atrial fibrillation with direct current cardioversion. <https://clinicaltrials.gov/ct2/show/NCT04300270>. Accessed 13 Dec 2021
 61. Parikh RB, Teeple S, Navathe AS (2019) Addressing bias in artificial intelligence in health care. *JAMA* 322:2377–2378
 62. Larson J, Mattu S, Kirchner L, Angwin J (2016) How we analyzed the COMPAS recidivism algo-

- rithm. <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Accessed 20 Dec 2021
63. Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366:447–453
64. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, Kohane IS, Saria S (2021) The clinician and dataset shift in artificial intelligence. *N Engl J Med* 385:283–286
65. Yu K-H, Beam AL, Kohane IS (2018) Artificial intelligence in healthcare. *Nat Biomed Eng* 2:719–731

Erkennung, Vorhersage und Behandlung von Vorhofflimmern mithilfe künstlicher Intelligenz

In diesem Beitrag wird der aktuelle Stand von Algorithmen des maschinellen Lernens zur Erkennung, Vorhersage und Behandlung von Vorhofflimmern zusammengefasst, zudem werden die Entwicklung und Prüfung von künstlicher Intelligenz in der Kardiologie und anderen Bereichen dargelegt. Nach heutigem Stand lässt sich Vorhofflimmern mithilfe künstlicher Intelligenz in 12-Kanal- oder 1-Kanal-Elektrokardiogrammen bzw. in Photoplethysmogrammen zuverlässig erkennen. Die Vorhersage von paroxysmalem oder neu auftretendem Vorhofflimmern hat die für den klinischen Einsatz erforderliche Genauigkeit noch nicht erreicht. Weitere Studien sind notwendig, um zu untersuchen, ob auf Basis des maschinellen Lernens eine Patientenselektion für Interventionen möglich ist.

Schlüsselwörter

VHF · KI · Maschinelles Lernen · Neuronale Netze · Disease Management · Mehrschichtiges Lernen



Ich bin Rhythmologin, weil ...



„...es für mich die perfekte Kombination aus Theorie und Praxis ist: Man kann sich vielen komplexen Fragestellungen widmen und gleichzeitig einer ebenfalls anspruchsvollen manuellen Tätigkeit nachgehen. Insbesondere die invasive Elektrophysiologie ist eines der abwechslungsreichsten Fächer, das neben manuellen Tätigkeiten wie Ablationen, Deviceimplantation und Occludertechnologien auch viele Tätigkeitsoptionen, Interaktionen und Forschungsoptionen mit anderen Bereichen der Kardiologie bietet. Dieser interdisziplinäre Charakter des Faches machen es für mich zu einem der spannendsten und innovativsten in der Inneren Medizin.“

Priv.-Doz. Dr. med. Arian Sultan,
Abteilung für Elektrophysiologie, Herzzentrum der Uniklinik Köln