

Applications of Machine Learning in Cardiac Electrophysiology

Rahul G Muthalaly and Robert M Evans

Monash Health, Melbourne, Australia

Abstract

Artificial intelligence through machine learning (ML) methods is becoming prevalent throughout the world, with increasing adoption in healthcare. Improvements in technology have allowed early applications of machine learning to assist physician efficiency and diagnostic accuracy. In electrophysiology, ML has applications for use in every stage of patient care. However, its use is still in infancy. This article will introduce the potential of ML, before discussing the concept of big data and its pitfalls. The authors review some common ML methods including supervised and unsupervised learning, then examine applications in cardiac electrophysiology. This will focus on surface electrocardiography, intracardiac mapping and cardiac implantable electronic devices. Finally, the article concludes with an overview of how ML may impact on electrophysiology in the future.

Keywords

Machine learning, artificial intelligence, surface electrocardiography, ablation, big data, neural network, cardiac devices

Disclosure: The authors have no conflict of interests to declare.

Received: 11 December 2019 **Accepted:** 29 April 2020 **Citation:** *Arrhythmia & Electrophysiology Review* 2020;9(2):71–7.

DOI: <https://doi.org/10.15420/aer.2019.19>

Correspondence: Rahul G Muthalaly, Monash Medical Centre, 246 Clayton Rd, Melbourne, VIC 3168, Australia. E: rahul.muthalaly@gmail.com

Open Access: This work is open access under the CC-BY-NC 4.0 License which allows users to copy, redistribute and make derivative works for non-commercial purposes, provided the original work is cited correctly.

Artificial intelligence (AI) has recently become a popular term in the technological world. AI refers to the simulation of human intelligence with the capacity for achieving goals within computers. Machine learning (ML) – a subtype of AI – refers to a statistical model that is able to independently learn to make inferences on new data based on data it has previously analysed. ML has markedly improved efficiency in multiple analytic domains including voice recognition, handwriting recognition, targeted marketing and robotics.¹ ML is a broad term that includes a family of methods extending from decision tree models to neural networks. Each method has different attributes that relate to their suitability for a given task. Interestingly, despite their recent popularity, neural networks have been in existence since at least the 1950s, when Marvin Minsky used an artificial neural network to solve a maze.² However, only in recent times has computing power sufficiently improved to allow the wider application of computationally intensive ML methods.

Aside from computing power, the other major development extending the reach of ML has been data acquisition and storage – an essential component to ML. Big data is a term used to describe the increasingly large and more complex datasets that form the basis for ML models.

Modern times have seen increasing accessibility and use of large volumes of data. This is apparent in medicine with the advent of electronic health records (EHRs), implantable electronic devices and advanced wearable monitors, all of which record unprecedentedly large volumes of biological data every second.³ ML has already contributed to healthcare advances in a number of specialties. This has taken various forms from assisting imaging and pathological diagnoses, to identifying novel disease risk factor

associations and phenotypic subgroups.^{4,5} Rajpurkar et al. used a convolutional neural network to develop a model that detected 14 different pathologies from a total of 420 frontal chest radiographs at a much faster rate than board-certified radiologists (1.5 minutes versus 240 minutes for all 420 radiographs). Their model was trained from a database of over 100,000 chest radiographs and could also anatomically localise where the pathologies (including pneumonia, effusions, masses and nodules) were present.⁶ As an example of unsupervised learning, Horiuchi et al. identified three phenotypic groups of acute heart failure patients that significantly predict mortality and re-hospitalisation.⁷ Within cardiac electrophysiology, near-term applications of novel data sources have begun appearing. Many of these are driven by industry–academic–healthcare partnerships, the most famous example of which is the Apple Heart study.⁸ However, significant late stage and outcomes data remain to be seen.

These early studies demonstrate some applications of ML in healthcare.⁹ Updating prior understanding and clinical practice developed over years will take time, but ML has the potential to improve disease definitions, classification and management.

Big Data

Data are a core requirement of ML methods. Theoretically, ML methods can be applied to datasets of any size. However, ML methods often benefit from large datasets as they provide more experience on which to train a model. 'Features' are components of a dataset that describe characteristics of the observations being studied. These features are fed into computational models that can then provide insight into the observations, for example clustering of similar observations into groups or prediction of outcomes. Large datasets used in this context are

referred to collectively as big data. There are many sources of big data in modern healthcare including EHRs, biobanks, clinical trial data and imaging datasets. The advent of ML methods has allowed enhanced insight into these resources.

The Three Vs of Big Data

There are three attributes that delineate the attributes of big data. The first V refers to volume of data. The larger the volume, the more features and experience a model can be trained and tested on. The second V refers to the velocity of data, which describes the speed at which the data is generated. With a higher velocity of data, models can remain more clinically relevant as they are retrained on current experience. The third V refers to variety, which reflects the diversity of the types of data contained in a dataset. Diverse data allows more features that can, in turn, increase generalisability and potentially the accuracy of the model.¹⁰ As a general rule, an increase in any of these attributes create a larger demand on the hardware and complexity of the method used to generate a model.

Potential Pitfalls of Big Data

The issues with big data can be related to the abovementioned Vs. High volume and velocity data requires significant computer processing power to analyse. Typically it is not an isolated volume or velocity issue, but a combination of the two together. This need for computational efficiency is one of the reasons novel ML methods outperform traditional statistical methods.

The extent of resources required to store and analyse data can be prohibitive. This limits the translation of data into clinical investigation and clinical practice. This is relevant in electrophysiology where cardiac devices create large volumes of data each second.

Another issue with big data is that it is often poorly organised and managed. This problem, also observed in small datasets, is amplified on a larger scale. Inaccurate data can be inappropriately included or not recognised for its limitations. Thus, the problem of inaccurate data producing inaccurate models remains.

Data sharing is an increasingly popular concept made easier with improvements in technological infrastructure.¹¹⁻¹³ Much of this is currently used for subgroup analyses, validation of prior work or the exploration of new hypotheses using trial data. Sharing of data will allow future researchers larger volumes of data to train and test models on. Many models exist for effective data sharing. Federated database systems allow geographically separate databases to be connected via networks. Thus, without merging the databases, researchers can submit queries to the federated database that interfaces with each individual database and provides results from all of them. This method allows the individual databases to remain heterogeneous and distinct. In contrast to this, distributed learning is a method that allows an ML algorithm to be trained using separate datasets. A central ML model is created and updated based on the training performed in each dataset. This model allows maintenance of data security and privacy whilst still harnessing the size of multiple datasets. However, opinions on data sharing as regards privacy and intellectual property, particularly in the medical industry sector, may remain a barrier to the spread of data sharing.¹⁴

Machine Learning Methods

ML algorithms use and require significant datasets in order to create models and test them. The foundation of ML is creation of a model on a

training dataset with subsequent validation of that model on an evaluation dataset (both of which are often subsets of an original large database). The model is then often run on a test dataset to provide an unbiased evaluation of its performance. The training dataset is used to fit the weights of a model, which detail the relationship between inputs and outcomes in a way specific to the chosen model. The validation dataset is then used to fine tune the model hyperparameters and evaluate the model. Once validated, models can be improved by being re-trained on new data. After re-training, models need to be validated again. These can increase the generalisability of the model across different populations/datasets. This allows the on-going improvement of the model with rapid responses to changing epidemiological or clinical patterns.

Machine Learning Computational Approaches/Algorithmic Principles

The overarching principle of ML is the use of training, evaluation and test datasets to create a valid model.¹⁵ ML methods are broadly categorised as supervised, unsupervised or reinforcement learning. Reinforcement learning aims to refine a strategy in a controlled environment stochastically. Reinforcement learning is outside the scope of this article.

Knowing which machine learning technique to apply is essential to achieving the objective, analogous to choosing an appropriate statistical technique in traditional methods or choosing an appropriate study design in epidemiology.

Supervised Learning Methods

Supervised learning (SL) algorithms aim to classify input data to the correct outputs based on prior input-output pairs that are correctly labelled. The need for labelling can be time consuming. However, the methods are very effective at classification using large datasets.

Examples of ML techniques used in SL include traditional linear and logistic regression, artificial neural networks (ANNs), support vector machines (SVMs), decision trees and random forests (RFs).^{15,16} Pictorial representations of common ML approaches are shown in *Figure 1*.

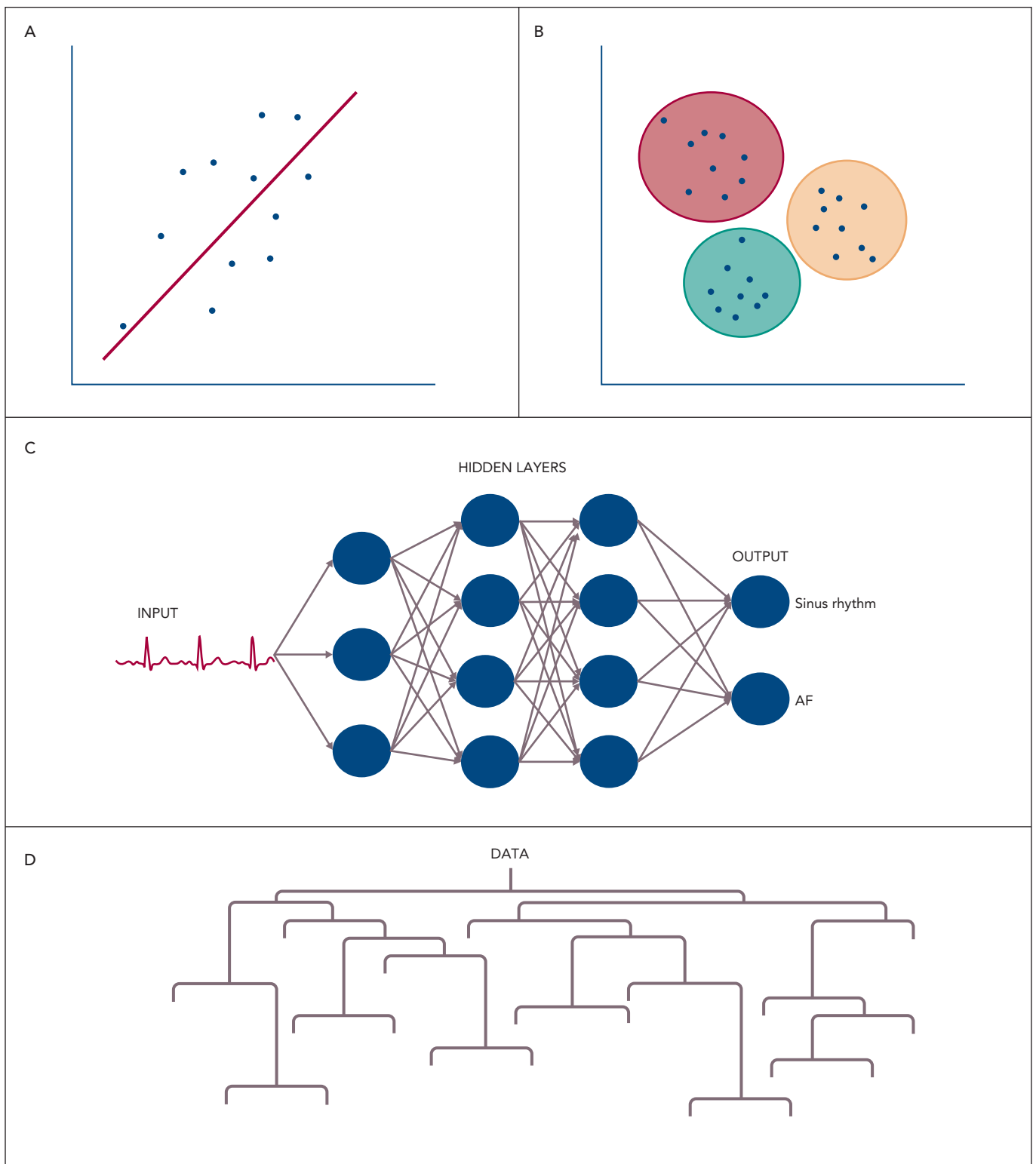
ANNs and SVMs are among the most common SL methods. They require intensive computational power and time and are prone to over fitting (where the model fits well for the training data but poorly for any new data). However, they are flexible regarding assumptions about the data and offer significant improvements in large datasets where traditional statistical methods struggle.¹⁷ ANNs can be layered and altered in order to offer more efficiency in dealing with complex and large data.

Decision trees and RFs are closely related. Decision trees are useful with smaller and simpler datasets. They classify data in a binary fashion along a chain of steps. Each branch following a step then has further classification at the next level. This then spreads out until an outcome is reached based on the series of steps taken to get there. The pictorial representation gives the shape to which decision trees are named.

RFs are an extrapolation of decision trees where multiple decision trees are combined and each tree is independently trained and verified. They also have the ability to exclude individual trees or components leaving them robust to outliers in meta-analysis and less prone to selection bias.

SL has its limitations including over fitting, the requirement of training and validation and the requirement for accurately labelled data, which

Figure 1: Pictorial Representations of Common Machine Learning Techniques



A: Linear regression where a line of best fit is estimated that allows future predictions to be made given new variables; B: Cluster analysis that separates observations into groups based on their similarity. In this case three groups have been determined; C: An artificial neural network demonstrating an input (ECG) that is fed through a number of connected nodes to match to an output (sinus rhythm); D: A decision tree diagram where values are determined at each branch until a final output is determined.

can be laborious.¹⁶ SL models can have poor generalisability and struggle transferring experiences to new data owing to over fitting whereby the model is very well fit only for the original dataset. As a result different models are often required for data from different populations. Each model requires large volumes of labelled data to train and then validate. Importantly, SL does not necessarily overcome prior

human biases in classification owing to the labelling step, which is often performed by humans and thus translates those biases into the model. This has had important implications in algorithms overweighting the importance of correlations as causation. This was the case with the Correctional Offender Management Profiling for Alternative Sanctions tool that analysed recidivism and was found to be using race as a

predictor in its algorithm.¹⁸ Each of these limitations result in varying degrees of bias, impacting on the data and future results.

Unsupervised Learning Methods

Unsupervised learning (UL) has a different goal to SL. In this method, the algorithm attempts to find patterns in the data without prior labelling. For example, in clustering the algorithm creates an unbiased set of categories without human intervention based on similarities between observations (e.g. creating groups of people based on their similarities). There are a variety of techniques included in UL, such as clustering, autoencoders and principal component analysis.¹⁵⁻¹⁷

UL is particularly useful when the dataset is very complex or if there is no natural fit to the data. In medicine, it is increasingly being used to identify disease phenotypes in heterogeneous conditions. This could suggest new classifications or stratifications that lead to more sophisticated treatment allocations. It can also be used in less established diseases, where data or literature are poorly developed or pathophysiology is not completely understood. As an example, Shah et al. clustered heart failure with preserved ejection fraction into phenotypic subtypes based on cardio-metabolic, cardio-renal and biochemical features.¹⁹

UL can produce multiple cluster schemes after analysis. Choosing which cluster scheme is most accurate or appropriate can be difficult and requires clinical interpretation. Importantly, just because a model defines clusters within a disease does not mean these clusters have any clinical relevance. For this to be successful sometimes it is necessary for UL to be combined with follow-up studies assessing the relevance and utility of these clusters. For example, Ahmad et al. identified differing mortality risks after clustering groups of patients with chronic systolic heart failure.²⁰

Machine Learning Applications in Cardiac Electrophysiology

ML is gradually expanding its utility in medicine, particularly in cardiology. Cardiac electrophysiology is particularly suitable for ML methods given its big data use and need for more accurate disease phenotype definitions and risk prediction. Returning to the three Vs, through implantable electronic devices, intracardiac mapping and wearable devices, high volume, velocity and variety data are produced. ML methods provide an opportunity to extract maximum clinical and public health benefit from this data.

Surface Electrocardiography

Surface ECGs can provide non-invasive, cheap and detailed information regarding arrhythmias. Importantly, because of these attributes, improving ECG interpretation could have substantial public health benefits.

There is a bulk of research involving ML and ECGs. While computer analysis of ECGs has been available on most machines for many years, these are often inaccurate because of the quality of ECG recorded. Traditional methods have relied on examining R-R intervals and P-wave presence. However, newer ML-based models have undertaken more sophisticated feature extraction to analyse a variety of rhythms.²¹

Research on ML in ECG began with basic improvements in pre-processing such as noise reduction or extraction of features such as P-wave or QRS complex characteristics.^{22,23} This initial work laid the foundation for future research to use these algorithms to create features that could be fed into classifiers such as ANNs and SVMs to further analyse ECGs. Much of the work performed in this area uses the MIT-BIH publicly available classified

data that consists of 25 ECG recordings that are 10 hours long.²⁴⁻²⁶ Importantly, this should herald caution in the assessment of diagnostic accuracy in such studies given the small sample sizes.

Many investigators have attempted to optimise the detection of AF using differing methods, duration of recordings and leads. Models differ by analysing either P-wave absence, R-R intervals or a combination of many ECG features. Ladavich and Ghoraani created a 'rate-independent' model that was able to identify AF using only a single cardiac cycle in order to overcome difficulties associated with detection of AF with rapid ventricular response. After feature extraction from the surface ECGs, an expectation-maximisation algorithm was used to train a model. The ultimate model had a sensitivity of 89.37% and specificity of 89.54% for detection of AF using just a single cardiac cycle.²⁴

Longer duration analysis has also been used to provide higher accuracy in the detection of AF. He et al. used a convolutional neural network on time-frequency features (as opposed to P-wave absence or R-R intervals) to train a model for AF detection.²⁶ The final model had an accuracy of 99.23%. Notably, the ECGs still required significant pre-processing before being classified using the convolutional neural network, which limits the potential for clinical use.

Other authors have used differing algorithms to train AF detection models. Kennedy et al. used RFs to train a model based on R-R intervals using their own database.²⁷ They later tested this model on the MIT-BIH database. The RF model had a specificity of 98.3% and sensitivity 92.8%, which is comparatively less than seen in many of the neural-network-based approaches. However, their training database was also considerably smaller and further direct comparisons of RF and ANN models are needed to determine their comparative accuracy in AF detection.

In contrast to ANNs and RFs, Asgari et al. used an SVM to train a model for detection of AF based on wavelet transformation.²⁸ The use of wavelet transformation obviates the need for pre-processing of P or R-waves that is required by many algorithms. However, this method still required feature extraction prior to classification with the SVM. The authors tested their model on the MIT-BIH database and found an area under the curve (AUC) of 0.995, which outperformed both a naïve Bayesian classifier and logistic regression based method.

There is less literature available on identification of ventricular arrhythmias. Mjihad et al. demonstrated that time-frequency analysis using a variety of models (logistic regression, ANN, SVM and bagging classifier) accurately identifies ventricular tachycardia (VT) and ventricular fibrillation (VF) on 12-lead ECG.²⁵ This method required pre-processing of the ECG signal for both noise reduction and computation of time windows. On testing, all models performed similarly, with ANNs having a 98.19% accuracy for VF and 98.87% accuracy for detection of VT.

Huang et al. demonstrated the ability of ANNs to localise atrioventricular accessory pathways in patients undergoing ablation.²⁹ Using features of delta wave polarity and R-wave duration as a proportion of the QRS complex, an ANN generated model trained on 90 cases correctly identified the site of accessory pathways in 58/60 test cases. In the two cases of misclassification, the accessory pathway was located in a contiguous region to the identified area.

The aforementioned studies have generally required significant pre-processing of the ECG signal prior to classification as well as involving

models trained for a specific arrhythmia. Hannun et al. created their own test and validation database.²¹ This included 91,232 single lead ECGs from the Zio patch-based electrode device. These ECGs were then fed to an ANN to create a model for classifying 12 different output rhythms (10 arrhythmias, sinus rhythm and noise). Importantly, their method did not require significant feature extraction or ECG signal pre-processing. The authors validated their findings against a committee of cardiologists who had classified the test dataset. The ANN achieved an AUC of 0.91 across all rhythms. Additionally, they compared their model to that of 'average' cardiologists outside the adjudication committee and found that the model outperformed cardiologists across every rhythm. This work represents a significant step forward in ECG classification as it demonstrates end-to-end machine learning, where raw data is inputted and diagnostic probabilities are outputted without the use of extensive data manipulation or pre-processing.

UL methods have, thus far, had more limited application in the analysis of surface ECG. Donoso et al. used a k-means clustering algorithm to separate AF on ECG into five different types based on frequency values. However, this work is yet to be validated by examination of the clinical significance of the five different types of AF.³⁰

AliveCor uses ML software built into their app to work in combination with an electrode band and phone or smartwatch.³¹ Their model is one of the first that has the ability to analyse the rhythm and diagnose AF in almost real time, as well as other ML models they have developed to identify long QT and hyperkalaemia off ECG.^{32,33} Twice weekly ECGs using their devices has been shown to be 3.9 times more likely to identify AF in high risk patients aged >65 years than routine monitoring.³¹ Additionally, recent work has examined the sensitivity of an AliveCor convolutional neural network using the Apple Watch heart rate, activity and ECG sensors as inputs compared to traditional implantable cardiac monitors for the detection of AF.³ The results suggest that the wearable monitor provides excellent sensitivity (97.5%) for detecting AF episodes lasting >1 hour but poor positive predictive value (39.9%). Given the excellent sensitivity, this strategy may help define those who would benefit from an implantable cardiac monitor post cryptogenic stroke.

Continual progress is being made in ML-based surface ECG analysis. With the advent of wearable technology, the availability of training data on which to improve models will increase and the quality of the raw data may also improve alongside the technology. However, these data will still require labelling for the implementation of SL methods, which represents a major resource barrier. External validation of the abovementioned models on larger datasets will be required before more sophisticated conclusions about the utility of such models can be made.

Intracardiac Mapping

ML is gradually emerging as a tool to improve the understanding and efficacy of ablation. Preclinical and early clinical work has focused on stratifying and classifying electrogram morphology to guide electrogram-based AF ablation.

In attempt to improve complex fractionated atrial electrogram (CFAE) based AF ablation, Schilling et al. created four classes of CFAEs based on patient data. They established classes that increased in complexity from class 0 to class 3.³⁴ They used a variety of features in order to create their complexity classes, including time domain descriptors, phase space descriptors, wavelet based descriptors, similarity of

active segments and amplitude based-descriptors. Once defined, they employed a fuzzy-decision tree (similar in methodology to a standard decision tree) to classify electrograms from 11 patients undergoing AF ablation. This algorithm had 81% accuracy for defining the CFAE class of an electrogram. The fuzzy decision tree model imparted the advantage of applying a probability that an electrogram belonged to a certain class unlike other algorithms. Duque et al. later validated and improved these CFAE classes by using a genetic algorithm (that optimised the included features) to further define the four classes.³⁵ After the initial consolidation of the classes, they used a k-nearest neighbour SL algorithm to classify electrograms. They demonstrated 92% accuracy for classification of CFAE class and performed simulation to demonstrate that the more complex CFAE classes were associated with rotor locations in a simulated model.

Orozco-Duque et al. similarly tried to improve classification of CFAEs using four features (two time-domain morphology based and two non-linear dynamic based) to separate four classes of fractionation.³⁶ They used a semi-supervised clustering algorithm to validate these classes on partially labelled data. They then went on to apply those classes to an unlabelled dataset and created clusters with reasonable separation. This work suggests that ML may be able to separate subtle differences in recorded electrograms and possibly provide feedback about areas most likely to result in AF termination after ablation.

McGillivray et al. used an RF classifier to identify re-entrant drivers of AF in a simulated model where the 'true' location of the re-entrant circuits were known.³⁷ This model used electrogram features to assess the rhythm, predict the location and reassess the predictions until focused on the source of the driver. The model correctly identified 95.4% of drivers whether one or more were present in the simulation.

Muffoletto et al. simulated AF ablation in a 2D model of atrial tissue using three different methods – pulmonary vein isolation, fibrosis-based and rotor-based ablation.³⁸ They were able to model the outcomes of these ablation strategies and use the outcome as labelled data for a ANN. Using different patterns in their AF simulation, the model was able to identify the successful ablation strategy in 79% of the simulations. This work serves as a proof-of-concept for ML prediction of optimal ablation strategies, but the difficulty of identifying the 'correct' ablation strategy amongst a number of options *in vivo* makes the possibility of clinical translation daunting. Computer modelling of AF is becoming more complex with the advancement of cell-level, tissue-level and organ-level models.^{39,40} These models have provided crucial insights into the relationships driving AF including the nature of automatic and rotational foci, the role of fibrosis and the effect of channel-types. ML may prove a useful tool in integrating the significant high-dimensional data produced by computer modelling. Possibilities include the definition of phenotypes using UL or the prediction of AF rotors based on patient-specific attributes, thus guiding ablation strategy using SL.

In an abstract presented at the American Heart Association 2018 scientific meeting, Alhousseini et al. described the use of a convolutional neural network to identify organised AF drivers that acutely terminate with ablation in persistent AF based on features extracted from spatial phase maps.⁴¹ They reported 95.2% accuracy for determining organised sites of activity that terminate AF upon ablation. These early results suggest another possibility for real-time ML-guided ablation strategies. However, the long-term efficacy of such an approach remains unclear.

ML is gradually entering the field of ablation and has the opportunity to integrate the significant volume of data generated by electroanatomical mapping systems and guide electrophysiologists to appropriate sites and methods of ablation. The use of SL in this way will face challenges given the significant data load and difficulties of creating labelled datasets, though the ultimate goal of machine-guided ablation may be achieved. However, UL methods may provide more insight into the patterns of arrhythmia seen during procedures.

Cardiac Implantable Electronic Devices

Cardiac implantable electronic devices (CIEDs) are an ideal target for ML methods given the high volume and velocity of data they produce. ML applications have allowed for risk stratification, improved arrhythmia localisation and streamlined remote monitoring which may significantly reduce the workload faced by electrophysiologists.⁴²⁻⁴⁴

CRT is an effective component of heart failure management in selected patients.⁴⁵ Benefit is restricted to those who meet current guideline criteria based on trial data. ML models may provide a more sophisticated method of identifying those likely to respond to CRT. Kalscheur et al. used multiple SL methods on data from the Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure (COMPANION) trial to construct a model capable of predicting outcomes with CRT.⁴² They employed an RF, decision tree, naïve Bayesian classifier and SVM with 48 features available in the clinical trial. They found that the RF model produced the best results, demonstrating an AUC of 0.74 for predicting outcomes. The quartile of highest risk predicted by the model had an eightfold difference in survival in comparison to the lowest risk quartile. When using traditional factors of QRS duration and morphology there was no significant association with outcomes. The features used in their model are easily clinically available. Cikes et al. employed UL for a similar purpose, seeking to identify high- and low-risk phenotypes for those likely to respond to CRT.⁴⁶ Using data from the Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy (MADIT-CRT) trial, they employed a k-means clustering algorithm to create groups with similar characteristics based on echocardiographic and clinical parameters. Of the four phenotypes they identified, two were associated with a better effect of CRT on heart-failure-free survival (HR 0.35 and HR 0.36 compared to non-significant). The advantage of using UL in this instance was the ability to identify the phenotypic aspects associated with CRT benefit. This is often not possible in SL owing to the nature of how the algorithms create associations, referred to as the 'black-box' problem.

Recognising the potential clinical difficulty of employing such models, Feeny et al. attempted to predict improved left ventricular ejection fraction with CRT using only nine features selected to optimise model performance.⁴⁷ These nine variables were employed in multiple ML models (including an SVM, RF, logistic regression, an adaptive boosting algorithm and a naïve Bayesian classifier). They found that the Bayesian classifier performed best demonstrating an AUC of 0.70, which was significantly better than guideline-based prediction. To demonstrate the clinical adaptability of their approach they generated a publicly available online calculator. The above studies are clinically relevant examples of using both UL and SL to risk stratify patients and help with decision making around significant interventions. The benefits of UL as regards transparency are observed, as well as the potential difficulties around clinical adaptation of complex models.

Rosier et al. examined the potential for ML to automate monitoring of alerts from CIEDs.⁴⁴ They used natural language processing to examine EHR data to determine the significance of AF alerts from CIEDs. The natural language processing algorithm was able to calculate CHA₂DS₂-VASc scores and anticoagulant status for each patient and thus, classify the importance of the AF alert. Their algorithm correctly stratified CHA₂DS₂-VASc scores of 0, 1 and ≥ 2 97% of the time. As a result of this, 98% of AF alerts were correctly classified with regard to their importance and the remaining misclassified alerts were overclassified, allowing for human review. This study highlights the ability of ML to act as an assistant to electrophysiologists by guiding attention to where it is needed.

In a similar vein of aiding electrophysiological intervention, Sanromán-Junquera et al. used ICD electrograms to localise exit sites during pace mapping for VT ablation.⁴³ Using implanted RV leads as sensors, they employed multiple algorithms including ANNs, SVMs and regression methods to identify left ventricular exit sites. Their SVM model produced the best results, localising the exit site to one eighth of the heart 31.6% of the time (where a random model would produce results of 12.5%). This work serves as a proof of concept for ML based ablation localisation, especially where a 12-lead ECG of the tachycardia may not be available. However, significant improvements are required before such methods are clinically applicable.

In an application of RF models, Shakibfar et al. used only ICD data without clinical variables to predict risk of electrical storm.⁴⁸ They developed 37 ICD electrogram-based features found during the four consecutive days prior to the onset of electrical storm (defined using device detection). They found that their RF model had an AUC of 0.80 for predicting electrical storm. The most relevant features were percentage of pacing and reduced daytime activity. These results are limited by the use of device-defined ventricular arrhythmias, however, they signify promise for the increasing use of the high dimensional data produced by CIEDs.

Conclusion

ML in electrophysiology is nascent. However, early work suggests the potential use that ML may have in stratification, diagnosis and therapy for arrhythmia. These methods may also affect the nature of the electrophysiologist's role into the future with increasing data sources and methods of analysis to add to each patient's data profile. As data derived from the use of wearable devices increases, consideration needs to be given to how these technologies will be implemented to ensure patient safety and appropriate use. In particular, the risk of over-diagnosis will need to be considered as wearable monitors become more commonplace. ■

Clinical Perspective

- Big data, easily collected from electronic health records and cardiac devices, allows the use of sophisticated models to generate new insights.
- Machine learning models can accurately diagnose multiple rhythms from short segments of surface electrocardiographs in almost real time.
- Optimisation of intra-cardiac mapping and implantable device analysis are areas that can significantly gain from increased machine learning integration owing to the large volume of data created in these fields.

1. Emerj. Everyday examples of artificial intelligence and machine learning. 2020. <https://emerj.com/ai-sector-overviews/everyday-examples-of-ai/> (accessed 15 May 2020).
2. Mc Kevitt P. *Daniel Crevier, Al: The tumultuous history of the search for artificial intelligence*. London and New York: Basic Books, 1993. <https://doi.org/10.1017/S0007087496302963>.
3. Wasserlauf J, You C, Patel R, et al. Smartwatch performance for the detection and quantification of atrial fibrillation. *Circ Arrhythm Electrophysiol* 2019;12:e006834. <https://doi.org/10.1161/CIRCEP.118.006834>; PMID: 31113234.
4. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol* 2019;20:e253–61. [https://doi.org/10.1016/S1470-2045\(19\)30154-8](https://doi.org/10.1016/S1470-2045(19)30154-8); PMID: 31044723.
5. Lakhani P, Prater AB, Hutson RK, et al. Machine learning in radiology: applications beyond image interpretation. *J Am Coll Radiol* 2018;15:350–9. <https://doi.org/10.1016/j.jacr.2017.09.044>; PMID: 29158061.
6. Rajpurkar P, Irvin J, Ball R, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med* 2018;15:e1002686. <https://doi.org/10.1371/journal.pmed.1002686>; PMID: 30457988.
7. Horiuchi Y, Tanimoto S, Latif A, et al. Identifying novel phenotypes of acute heart failure using cluster analysis of clinical variables. *Int J Cardiol* 2018;262:57–63. <https://doi.org/10.1016/j.ijcard.2018.03.098>; PMID: 29622508.
8. Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. *N Engl J Med* 2019;381:1909–17. <https://doi.org/10.1056/NEJMoa1901183>; PMID: 31722151.
9. Krittanawong C. Future physicians in the era of precision cardiovascular medicine. *Circulation* 2017;136:1572–4. <https://doi.org/10.1161/CIRCULATIONAHA.117.029676>; PMID: 29061572.
10. Andreu-Perez J, Poon CC, Merrifield RD, et al. Big data for health. *IEEE J Biomed Health Inform* 2015;19:1193–208. <https://doi.org/10.1109/JBHI.2015.2450362>; PMID: 26173222.
11. Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials – a requirement of the International Committee of Medical Journal Editors. *N Engl J Med* 2017;376:2277–9. <https://doi.org/10.1056/NEJMe1705439>; PMID: 28581902.
12. Kiley R, Peatfield T, Hansen J, Reddington F. Data sharing from clinical trials – a research funder's perspective. *N Engl J Med* 2017;377:1990–2. <https://doi.org/10.1056/NEJMs1708278>; PMID: 29141170.
13. Pencina MJ, Louzao DM, McCourt BJ, et al. Supporting open access to clinical trial data for researchers: the Duke Clinical Research Institute–Bristol-Myers Squibb Supporting Open Access to Researchers Initiative. *Am Heart J* 2016;172:64–9. <https://doi.org/10.1016/j.ahj.2015.11.002>; PMID: 26856217.
14. Naudet F, Sakarovich C, Janiaud P, et al. Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies published in the BMJ and PLOS Medicine. *BMJ* 2018;360:k400. <https://doi.org/10.1136/bmj.k400>; PMID: 29440066.
15. Russell S, Norvig P. *Artificial Intelligence: A Modern Approach*. 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2009.
16. Mitchell TM. *Machine Learning*. New York: McGraw-Hill, 1997.
17. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge, MA: MIT Press, 2016.
18. ProPublica. Machine bias. 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (accessed 15 May 2020).
19. Shah SJ. Precision medicine for heart failure with preserved ejection fraction: an overview. *J Cardiovasc Transl Res* 2017;10:233–44. <https://doi.org/10.1007/s12265-017-9756-y>; PMID: 28585183.
20. Ahmad T, Pencina MJ, Schulte PJ, et al. Clinical implications of chronic heart failure phenotypes defined by cluster analysis. *J Am Coll Cardiol* 2014;64:1765–74. <https://doi.org/10.1016/j.jacc.2014.07.979>; PMID: 25443696.
21. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019;25:65–9. <https://doi.org/10.1038/s41591-018-0268-3>; PMID: 30617320.
22. Ochoa-Brust A, Mena L, Felix V. Noise-tolerant neural network approach for electrocardiogram signal classification. *Proceedings of the International Conference on Compute and Data Analysis 2017:277–82*. <https://doi.org/10.1145/3093241.3093269>.
23. Pourbabae B, Lucas C. Automatic detection and prediction of paroxysmal atrial fibrillation based on analyzing ECG signal feature classification methods. Presented at Cairo International Biomedical Engineering Conference 2008, Cairo, Egypt, 18–20 December 2008. <https://doi.org/10.1109/CIBEC.2008.4786068>.
24. Ladavich S, Ghoraani B. Rate-independent detection of atrial fibrillation by statistical modeling of atrial activity. *Biomed Signal Process Control* 2015;18:274–81. <https://doi.org/10.1016/j.bspc.2015.01.007>.
25. Mjahad A, Rosado-Munoz A, Batailler-Mompean M, et al. Ventricular fibrillation and tachycardia detection from surface ECG using time-frequency representation images as input dataset for machine learning. *Comput Methods Programs Biomed* 2017;141:119–27. <https://doi.org/10.1016/j.cmpb.2017.02.010>; PMID: 28241963.
26. He R, Wang K, Zhao N, et al. Automatic detection of atrial fibrillation based on continuous wavelet transform and 2D convolutional neural networks. *Front Physiol* 2018;9:1206. <https://doi.org/10.3389/fphys.2018.01206>; PMID: 30214416.
27. Kennedy A, Finlay DD, Guldenring D, et al. Automated detection of atrial fibrillation using R-R intervals and multivariate-based classification. *J Electrocardiol* 2016;49:871–6. <https://doi.org/10.1016/j.jelectrocard.2016.07.033>; PMID: 27717571.
28. Asgari S, Mehriani A, Moussavi M. Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine. *Comput Biol Med* 2015;60:132–42. <https://doi.org/10.1016/j.combiomed.2015.03.005>; PMID: 25817534.
29. Huang D, Yamauchi K, Inden Y, et al. Use of an artificial neural network to localize accessory pathways of Wolff-Parkinson-White syndrome with 12-lead electrocardiogram. *Med Inform Internet Med* 2005;30:277–86. <https://doi.org/10.1080/14639230500367670>; PMID: 16531354.
30. Donoso FI, Figueroa RL, Lecannelier EA, et al. Clustering of atrial fibrillation based on surface ECG measurements. *Conf Proc IEEE Eng Med Biol Soc* 2013;2013:4203–6. <https://doi.org/10.1109/EMBC.2013.6610472>; PMID: 24110659.
31. Halcox JPI, Wareham K, Cardew A, et al. Assessment of remote heart rhythm sampling using the Alivacor heart monitor to screen for atrial fibrillation: the REHEARSE-AF Study. *Circulation* 2017;136:1784–94. <https://doi.org/10.1161/CIRCULATIONAHA.117.030583>; PMID: 28851729.
32. Karacan M, Celik N, Gul EE, et al. Validation of a smartphone-based electrocardiography in the screening of QT intervals in children. *Northern Clin Istanbul* 2019;6:48–52. <https://doi.org/10.14744/mci.2018.44452>; PMID: 31180383.
33. Galloway CD, Valys AV, Shrelbati JB, et al. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. *JAMA Cardiol* 2019;4:428–36. <https://doi.org/10.1001/jamacardio.2019.0640>; PMID: 30942845.
34. Schilling C, Keller M, Scherr D, et al. Fuzzy decision tree to classify complex fractionated atrial electrograms. *Biomed Tech (Berl)* 2015;60:245–55. <https://doi.org/10.1515/bmt-2014-0110>; PMID: 25781659.
35. Duque SJ, Orozco-Duque A, Kremen V, et al. Feature subset selection and classification of intracardiac electrograms during atrial fibrillation. *Biomedical Signal Process Control* 2017;38:182–90. <https://doi.org/10.1016/j.bspc.2017.06.005>.
36. Orozco-Duque A, Bustamante J, Castellanos-Dominguez G. Semi-supervised clustering of fractionated electrograms for electroanatomical atrial mapping. *Biomed Eng Online* 2016;15:44. <https://doi.org/10.1186/s12938-016-0154-5>; PMID: 27167088.
37. McGillivray MF, Cheng W, Peters NS, et al. Machine learning methods for locating re-entrant drivers from electrograms in a model of atrial fibrillation. *R Soc Open Sci* 2018;5:172434. <https://doi.org/10.1098/rsos.172434>; PMID: 29765687.
38. Muffoletto M, Fu X, Roy A, et al. Development of a deep learning method to predict optimal ablation patterns for atrial fibrillation. Presented at IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) 2019, Siena, Italy, 9–11 July 2019. <https://doi.org/10.1109/CIBCB.2019.8791475>.
39. Aronis KN, Ali RL, Liang JA, et al. Understanding AF mechanisms through computational modelling and simulations. *Arrhythm Electrophysiol Rev* 2019;8:210–9. <https://doi.org/10.15420/aer.2019.28.2>; PMID: 31463059.
40. Ciaccio EJ, Wan EY, Saluja DS, et al. Addressing challenges of quantitative methodologies and event interpretation in the study of atrial fibrillation. *Comput Methods Programs Biomed* 2019;178:113–22. <https://doi.org/10.1016/j.cmpb.2019.06.017>; PMID: 31416540.
41. Alhusssein M, Abuzaid F, Swerdlow M, et al. Sites where ablation terminated atrial fibrillation identified by machine learning models. Presented at American Heart Association Scientific Sessions 2018, Chicago, IL, 10–12 November 2018. Abstract 13161.
42. Kalscheur MM, Kipp RT, Tattersall MC, et al. Machine learning algorithm predicts cardiac resynchronization therapy outcomes. *Circ Arrhythm Electrophysiol* 2018;11:e005499. <https://doi.org/10.1161/CIRCEP.117.005499>; PMID: 29326129.
43. Sanromán-Junquera M, Mora-Jiménez I, Almendral J, et al. Automatic supporting system for regionalization of ventricular tachycardia exit site in implantable defibrillators. *PLoS One* 2015;10:e0124514. <https://doi.org/10.1371/journal.pone.0124514>; PMID: 25910170.
44. Rosier A, Mabo P, Temal L, et al. Personalized and automated remote monitoring of atrial fibrillation. *Europace* 2016;18:347–52. <https://doi.org/10.1093/europace/euv234>; PMID: 26487670.
45. Bristow MR, Saxon LA, Boehmer J, et al. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. *N Engl J Med* 2004;350:2140–50. <https://doi.org/10.1056/NEJMoa032423>; PMID: 15152059.
46. Cikes M, Sanchez-Martinez S, Claggett B, et al. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *Eur J Heart Fail* 2019;21:74–85. <https://doi.org/10.1002/ejhf.1333>; PMID: 30328654.
47. Feeny AK, Rickard J, Patel D, et al. Machine learning prediction of response to cardiac resynchronization therapy: Improvement versus current guidelines. *Circ Arrhythm Electrophysiol* 2019;12:e007316. <https://doi.org/10.1161/CIRCEP.119.007316>; PMID: 31216884.
48. Shakibfar S, Krause O, Lund-Andersen C, et al. Predicting electrical storms by remote monitoring of implantable cardioverter-defibrillator patients using machine learning. *Europace* 2019;21:268–74. <https://doi.org/10.1093/europace/euy257>; PMID: 30508072.