Artificial Intelligence in Cardiac Imaging

Karthik Seetharam, MD, Sirish Shrestha, MSc, and Partho P Sengupta, MD, DM

Division of Cardiology, West Virginia University Heart and Vascular Institute, Morgantown, WV

Abstract
Machine learning (ML), a subset of artificial intelligence, is showing promising results in cardiology, especially in cardiac imaging. ML algorithms are allowing cardiologists to explore new opportunities and make discoveries not seen with conventional approaches. This offers new opportunities to enhance patient care and open new gateways in medical decision-making. This review highlights the role of ML in cardiac imaging for precision phenotyping and prognostication of cardiac disorders.

Keywords
Machine learning, artificial intelligence, cardiac imaging, precision phenotyping, deep learning, convolutional neural network

Numerous advances in cardiovascular imaging technology paralleled by the simultaneous growth in wearable technology, mobile health devices, and electronic medical record integration have encouraged increasingly complex and large multi-dimensional data acquisition.1 The influx of data with each scan is exponentially rising in all cardiovascular imaging modalities and will exceed the capabilities of current statistical software.2 Artificial intelligence (AI) has sparked remarkable progress in various aspects of technology from speech recognition to automated driving.3,4 Machine learning (ML), a subset of AI, can harvest information from this vast data matrix to improve disease prognostications and survival prediction.5 The current model of image acquisition and interpretation has led to a number of issues in timing, efficiency, and inaccurate diagnosis.5 ML can help transcend the gap between the rapid growth of cardiac imaging and clinical care by improving image acquisition, interpretation, subsequent decision-making, and reducing costs.6 Furthermore, ML facilitates a number of opportunities for data-driven discoveries and innovation not typically seen with conventional statistical approaches.7 In this article, we review several contributions that ML can make in cardiac imaging.

Types of Machine Learning
ML algorithms strive to learn and find natural patterns in data to support decisions, automation, and risk mitigation to enable generalizability.8,9 It generally results in better performance if the algorithm learns from large heterogeneous datasets for prediction from an unlearnt dataset, such as one that has not been subjected to ML. It aspires to learn the data in multiple different ways that can be broadly subdivided into supervised, unsupervised, semi-supervised, and reinforcement learning (Table 1 and Figure 1). Recent investigations have shown the ability to use supervised and unsupervised ML for cardiac imaging.6,9 Supervised and unsupervised learning are the commonly used approaches.10 Supervised learning works with datasets with labeled variables or classified outcomes, where it is trained to build a model from a select feature derived from any imaging data sample and clinical variables along with the outcome of interest.11 It reacts to the feedback based on corresponding labels from modalities, such as ECG, CCT, and cardiac MRI (CMR). This subsequently contributes to prediction and risk stratification of cardiovascular diseases. In contrast to supervised learning, an unsupervised learning algorithm must decipher the data without labels or interventions.11 Semi-supervised learning is a type of learning process that has a mixture of labeled and unlabeled classes within the dataset.2 It has a role in speech and image recognition. Reinforcement learning uses reward criteria, like human psychology, learning through trial and error. Although reinforcement learning has had limited development in cardiology, there is a growing interest in these techniques in clinical and research settings.2 Physicians must remain constantly aware of their data to prevent bias creeping into the models.10 For example, sampling bias can creep in if the training data do not accurately represent the heterogeneity in the cardiovascular diseases. Prejudicial bias may occur if the sample is affected by cultural, ethnic, or gender factors. Measurement bias can also occur. It can occur if the model is trained on incorrect and noisy image data or inaccurate measurements from various modalities.
Deep learning is a subset of ML and has sparked much interest in the health sector.\textsuperscript{10} Deep learning uses several layers similar to neuronal architecture in the human brain that enables reasoning and interpretation. Information is progressively processed through this neuronal-like hierarchy to analyze and interpret information. It learns through a series of iterations where it extrapolates predictive properties which are superior to conventional ML algorithms.\textsuperscript{10} Unlike other ML algorithms, the performance of deep learning improves with larger datasets.\textsuperscript{5} Deep learning is becoming increasingly popular due to paralleled growth in cloud infrastructures and improving computer performance.\textsuperscript{7} It is frequently used in image segmentation, and uses associations based on previous experience to increase the chances of correct classification by the ML algorithm. There is great potential for deep learning to have significant impact on cardiovascular imaging in the future.\textsuperscript{11}

There are algorithms within deep learning. Convolutional neural network (CNN) algorithms are the most frequently used.\textsuperscript{12} CNN algorithms consist of a convolutional part and a fully connected part. The convolutional part enables feature extraction to occur and the fully connected part allows classification or regression.\textsuperscript{12} The convolutional part allows the generation of feature maps based on the parameters used in the analysis. The commonly used Googlenet and ResNet are based on CNN-like architectures.\textsuperscript{12}

### Machine Learning in Echocardiography

Echocardiography frequently serves as the first line of diagnostic imaging and is an integral part of cardiology practice.\textsuperscript{1} There are numerous echocardiographic variables, and this, along with the development of vector flow mapping and speckle tracking have overwhelmed a clinician’s ability to properly evaluate results.\textsuperscript{13} ML can use the vast amount of information present in echocardiography to uncover valuable information and has seen exponential growth in the use of various algorithms in different modalities in cardiac imaging research (Figure 2).\textsuperscript{7}

Most studies have emphasized the use of a supervised ML framework for developing diagnostic predictions. For example, Zhang et al. used a CNN for automatic interpretation of ECG in 14,035 ECGs over a span of 10 years.\textsuperscript{14} The authors showed that the ML measurements were superior to manual measurements across numerous standards, such as the correlation of left atrial and ventricular volumes. Some aspects of this study must be interpreted with caution and require further validation. The algorithm may have correlated well with manual measurements, but there were wider limits.\textsuperscript{15} This emphasizes the variability of ECG findings in daily practice, and the lack of gold standard metrics, such as CMR, cannot be used for comparison.

Supervised ML has also been used for predicting future adverse cardiac events using echocardiography data. Samad et al. used the random forest ML algorithm to predict survival after ECG in a population of 171,510 patients.\textsuperscript{22} The model was compared with the logistic regression model using three different inputs involving a variety of ECG and clinical parameters by mean area under the curve (AUC). The random forest algorithm achieved superior prediction accuracy (all AUC >0.82) against common clinical risk scores (AUC 0.69–0.79). Additionally, it outclassed logistic regression models (p<0.001 and all survival durations). However, a unique feature demonstrated by Samad et al. was pursuing a broad initial hypothesis, rather than contemporary hypothesis-driven research.\textsuperscript{27} This can help overcome gaps in knowledge by revising initial inquiry or by leading to new questions.\textsuperscript{16}

Similarly, Madani et al. applied a CNN algorithm to 267 transthoracic ECGs with 15 standard views to showcase real-life variation.\textsuperscript{17} The overall test accuracy was 97.8% for the ML model across 12 views. Among low-resolution images, there was a 91.7% accuracy for the ML algorithm in relation to 70.2–84.0% for board certified echocardiographers. In another report by Madani et al., they used deep learning classifiers for automatic interpretation of ECGs.\textsuperscript{27} Madani et al. obtained an accuracy of 94.4% for 15 echocardiographic view classifications of still images and 91.2% accuracy for binary left ventricular (LV) hypertrophy view classification. Afterwards, the researchers used a semi-supervised generative adversarial network model that showed 80% accuracy in view classification and 92.3% accuracy for LV hypertrophy.\textsuperscript{25}

### Machine Learning in Nuclear Cardiology

The potential of ML in nuclear cardiology is vast and can greatly shape decision-making for cardiologists.\textsuperscript{28} Arsanjani et al. investigated the accuracy of myocardial perfusion imaging (MPI) for coronary artery disease (CAD) prediction by using a supervised learning algorithm in 957 studies using perfusion and functional variables.\textsuperscript{26} The outcomes of the ML algorithm were compared with automatic quantification software and two experienced readers. The sensitivity and specificity of ML were superior (p<0.005) to the quantification software and two experienced readers. The receiver operating curve (ROC) area under the curve for the ML algorithm (0.92) was statistically superior to both readers (0.87 and 0.88, p<0.03). Arsanjani et al. conducted another study to predict CAD by incorporating clinical and imaging data in 1,181 patients with ML algorithms to improve single-photon emission computed tomography (SPECT) accuracy.\textsuperscript{27} A supervised ML algorithm or LogitBoost was used. The ROC curve for the ML architecture (0.94 ± 0.01) was statistically superior to the total perfusion deficit and used two readers (p<0.001).
Haro Alonso et al. developed an ML algorithm to predict the risk of cardiac death derived from a combination of adenosine myocardial perfusion using myocardial perfusion SPECT (MPS) and clinical data in 8,321 patients and 551 cases of cardiac death. This was compared with logistic regression. Remarkably, the logistic regression was outperformed by all ML approaches (AUC 0.76; 14 features). The support vector machine demonstrated greatest accuracy (AUC 0.83; p <0.0001; 49 features). The ML algorithm was superior to logistic regression by providing the best AUC for showing the risk of cardiac death. The authors clearly demonstrated the superiority of ML models to improve prognostic value from multi-dimensional MPS and clinical variables while simultaneously increasing interpretability. However, this should not underscore our preference towards ML models over conventional statistical models, rather it changes our previous beliefs regarding the inability to integrate high-dimensional data and identify unique properties which may otherwise affect our perception of parameters.

Nonetheless, the study was retrospective and did not occur in real time.

Machine Learning in Cardiac CT
Motwani et al. investigated the use of an ML algorithm to predict 5-year mortality in CT scans in reference to conventional cardiac metrics in 10,030 patients with possible CAD. The ML algorithm demonstrated a statistically significant (p<0.001) higher AUC (0.79) than fractional flow reserve (0.61) or CT severity scores (segmental stenosis score 0.64, segment involvement score 0.64, Duke index 0.62) for predicting 5-year all-cause mortality. Although the study offers exciting opportunities, there are a number of hurdles to overcome to successfully implement these approaches in clinical practice, and institutions will require dedicated IT teams to manage the data that need to be anonymized. There are significant costs associated with management and regulation. Clinical trials may be required to determine the effectiveness of these various ML processes.

Rosandaelt al. explored the use of an ML algorithm in 8,844 patients to predict major cardiovascular events using only CT variables in evaluation to CT severity scores for patients with suspected CAD. Remarkably, the AUC for the ML algorithm (0.77) was far better than CT severity scores (0.69–0.70) with a statistical significance (p<0.001) for prediction of major cardiovascular events.

Zreik et al. used a CNN algorithm to automatically calculate fractional flow reserve from coronary CT angiography in 166 patients who underwent invasive coronary angiography. The area under the ROC was 0.74 (±0.02). When sensitivity levels measured 0.60, 0.70, and 0.80, the equivalent specificity was 0.77, 0.71, and 0.59.
Machine Learning in Cardiac MRI

The application of ML learning in CMR is quite limited and there is significant opportunity for growth. However, there has been some notable works that demonstrate the viability of ML in CMR. For example, Winther et al. used a deep learning ML algorithm for automatic segmentation of the right ventricular (RV) and LV endocardium and epicardium for assessing cardiac mass and function parameters from a variety of datasets. They found that learning architecture accomplished a comparable outcome in relation to human experts. Nevertheless, the findings must be taken with a degree of caution due to small sample sizes.

Bai et al. applied a fully convolutional network for automated analysis of CMR images from a large database consisting of 93,500 images in 5,000 patients for measuring left and RV mass and volumes. On the short axis image test of 600 patients, the Dice metric measured 0.94 for LV cavity, 0.88 for LV myocardium, and 0.90 for RV cavity. In addition, the average Dice metric measured 0.93 for the left atrial cavity in two chamber view, 0.95 for left atrial cavity in four chamber view, and 0.96 in right atrial cavity in two chamber view. Bai et al. demonstrated ML automated methods had values comparable with human experts.

Tan et al. assessed the function of a convolutional network, a supervised learning approach for automatic segmentation of the left ventricle in all short axis slices. It was applied to a number of publicly available datasets which included the LV segmentation challenge dataset consisting of 200 CMR imaging sets with unique cardiac pathology. They obtained a Jaccard index of 0.77 in the LV segmentation challenge dataset. Furthermore, they obtained a continuous ranked probability score of 0.0124 with the Kaggle Second Annual Data Science Bowl. Findings from Tan et al. showed the potential of the ML algorithm in automatic LV segmentation in CMR.

Machine Learning for Identifying Cardiac Phenotypes

The exponential rise in data size and complexity will make it difficult for physicians to analyze data in a clinically meaningful manner. In DNA analysis, phenomapping is used to comprehend vast quantities of data by subdividing subjects into various categories. Recently, ML algorithms have expanded the role of phenomapping in cardiac imaging.

Tabassian et al. explored the role of ML in assessing the timing and amplitude of the LV long-axis myocardial motion and deformation at stress and rest in 100 prospectively recruited patients by using unsupervised and supervised learning. This was compared with conventional measurements in patients with heart failure with preserved ejection fraction (HFpEF), healthy, breathless, and hypertensive people. The learnt strain rate parameters showed the highest accuracy for categorizing subjects into four groups (overall 57%; HFpEF patients 81%) and into two classes (asymptomatic versus symptomatic; AUC 0.89; accuracy 85%; sensitivity 86%, specificity 82%). When comparing ML with conventional measurement for strain, it demonstrated the highest improvement in accuracy for the two-class task (+23%, p<0.001), comparison with +11% (p<0.001) using velocity and +4% (p<0.05) using strain.

Tabassian et al. encountered an issue with their data, which consisted of 36 segmental curves for each patient with 208 and 123 time points at rest and exercise, respectively. They used unsupervised learning to break down the complexity of the data into reduced dimensions. It is an approach to represent data with fewer features, potentially to reduce noise and multicollinearity, from high-dimensional data obtained from imaging modalities and clinical data. Subsequently, they used a supervised classification algorithm to categorize patients into one of four...
These two regions were connected from the top and the bottom by moderate AS. The region of severe AS showed greatly elevated risk of balloon valvuloplasty and transcatheter aortic valve replacement. After surgical intervention, many patients were in the regions of mild and moderate AS. The results were further validated in mouse models with similar results. The authors showed that ML has the potential to facilitate precise recognition of phenotypic patterns of the left ventricle during AS. The model was able to show the movement of the patients from severe to mild pre- and post-aortic valve replacement (Figure 4). Further studies and clinical trials will be required to validate these findings.

**Role of Deep Learning in Cardiac Imaging**

Deep learning has unprecedented potential to revolutionize the field of cardiac imaging. Steady and progressive advances in computer processing capabilities and cloud infrastructure have propelled the growth of deep learning.2

Betancur et al. led a multicenter investigation assessing automatic prediction of CAD obstruction in 1,638 patients by MPI implemented through deep learning in relation to total perfusion deficit.23 These patients underwent stress Tc-Sestambi or tetrofosmin MPI, and invasive coronary artery angiography was done within 6 months. Only 1,018 (62%) had obstructive CAD among the total population. Deep-learning ML showed a higher AUC than total perfusion deficit for CAD prediction (per patient: 0.80 versus 0.78; per vessel 0.76 versus 0.73: p<0.01). If the ML matched the total perfusion deficit specificity, per-patient sensitivity (79.8% to 82.3%, p<0.05) and per-vessel (64.4% to 69.8%, p<0.01) sensitivity increased. Although Betancur et al. performed a remarkable feat, there are still issues that need to be addressed.24 The sensitivity could be higher for the algorithm. The relationship between epicardial anatomy and cellular perfusion is not ideal. As a result, the study did not focus on a definite endpoint, such as MI or death. In reality, it is also very difficult to define who is an expert and who is not.

Yang et al. explored the role of deep learning ML for left ventricle segmentation in publicly available datasets which included SATA-13 and LV-09.25 Sometimes accurate segmentation of the left ventricle from CMR can be challenging due to significant variation in intensity levels, structural shape issues, and respiratory movement artefacts. The deep-learning ML showed a 0.83 in averaged Dice metric on SATA-13 dataset and 0.95 averaged Dice metric for the LV-09 dataset. As a result, it compared satisfactorily with other automatic segmentation approaches.

**Issues to be Resolved Before Machine Learning is Widely Adopted**

Although the promise and early fruits of ML in cardiology have been encouraging, it is far from straightforward. In parallel with the rapid growth of ML, many steps must be taken to facilitate the transition of ML into clinical practice so that patient care can benefit.

A universal reference standard may need to be adopted by institutions to fully validate and assess the accuracy of ML learning approaches for each diagnostic modality.26 Depending on the standard chosen for evaluation, ML might show differing results. Furthermore, some form of a universal approach is mandatory for data standardization. There
Artificial Intelligence in Cardiac Imaging

are discrepancies between each institution. Each center has their own classification, protocol, and different acquisition protocols. If ML algorithms are universally accepted, there are subsequent difficulties in clinical implementation, and maintenance of data quality. Some form of homogenization of clinical data recording and imaging protocols may become necessary to serve as a common input for the use of ML. In addition, information from electronic health records needs to be streamlined with imaging databases.

Another prevailing issue in ML adoption is cost. For ML algorithms to thrive in any institution, significant investment is needed. A surrounding IT team and ecosystem may be needed to help institutions during ML training, and development. As ML technology continues to grow and evolve, more investments will be needed to keep up with trends and development.

Large data are pivotal for ML training, and accuracy increases proportionately. For any single institution, it can be difficult to obtain data large enough to reach any meaningful conclusion with any ML algorithm. As a result, some form of data sharing needs to be established between centers to improve ML accuracy. For successful data sharing, certain issues need to be resolved: the datasets need to be completely anonymized, and multiple institutional review board approvals are needed for the sharing of data among institutions and this can be laborious. If datasets are publicly available, this can significantly help the growth of ML.

Another important issue for the adoption of ML in academic centers is the ethics of AI. For discussion’s sake, AI has no moral compass. Unintentionally, AI can make errors in judgement when confronted with unusual scenarios. Its ‘black box’ nature is not easily understood. These properties can be an issue in the real world because of the flawed and imperfect nature of the world. Engineering cannot address every issue in the design of these algorithms. Key attributes need to be decided before using and designing any algorithm. One way of addressing these issues is to have AI classes introduced into medical school curricula to help trainee doctors to responsibly use and comprehend AI. Physicians need to constantly update themselves regarding advances in AI to fully tap its potential.

As ML continues to expand across academic institutions, there are potential benefits. This technology facilitates the possibility of reducing the clinical burden by mitigating repetitive activities, such as measurements, data preparation, or quality control. As a result, more dedicated time can be steered toward medical interpretation and complex decision-making. Furthermore, ML can supplement and aid the clinical acumen of physician by offering more possibilities. This can allow physicians to ask patients vital diagnostic questions which can foster improved patient care. These possibilities of ML point towards a healthier harmony between physician and patient in years to come.

Conclusion
ML has revolutionary potential to completely alter current clinical guidelines in diagnostic evaluation and decision-making. It provides
limitless possibilities for analyzing data. As data continue to become larger and more complex, ML will become pivotal. In this data-driven era, traditional beliefs in statistics must evolve with changing times and embrace data-driven discoveries. At the same time, cardiologists must evaluate the legal and ethical ramifications and develop a set of reference standards for the successful implementation of ML in clinical care.