Demystifying Deep Learning for Cardiovascular Imaging: A Review

Ramsey M. Wehbe, Aggelos K. Katsaggleos, Kristian J. Hammond, Ha Hong, Faraz S. Ahmad, David Ouyang, Sanjiv J. Shah, Patrick M. McCarthy, James D. Thomas

Manuscript Word count: 4,495

Tables/Figures: 6

Supplemental Tables/Figures: 7

Narrative Review
full info

Up-to-date review for clinicians on a topic of general common interest from the perspective of internationally recognized experts in these disciplines.

The focus should be an update on current understanding of the physiology of the disease or condition, diagnostic consideration, and treatment.

These reviews should address a specific question or issue that is relevant for clinical practice.

- **2000-3500 words**
- 50-75 references
- **≤5 tables and/or figures**
- 3-part structured abstract
- No Key Points
- Subtitle should be "A Review"

Abstract:

Importance:

Artificial intelligence (AI), driven by revolutionary advances in deep learning (DL), is transforming the field of cardiovascular imaging (CVI) at a breathtaking pace. While DL for cardiovascular imaging is still in its infancy, research is accelerating to aid in the acquisition, processing, and/or interpretation of CVI across various modalities, with several commercial products already in clinical use. It is imperative that cardiovascular imagers are familiar with DL systems, including a basic understanding of how they work, their relative strengths compared to other automated systems, and potential pitfalls in their implementation. The goal of this manuscript is to review the history,

methodology, and application of DL for CVI in a simple and digestible fashion towards demystifying this emerging technology.

Observations:

The current DL boom has been made possible by advances in mathematical algorithms, curation of massive digital datasets, and exponential growth in computational power. DL at its core is simply the application of a series of mathematical operations and non-linear transformations that translate input data into a desired output. Based on artificial neural networks that are inspired by the human nervous system, there are several types of DL architectures suited to different tasks – convolutional neural networks are particularly adept at extracting valuable features from CVI data. We discuss challenges in the development and implementation of DL systems, including avoiding overfitting, preventing systematic bias, improving explainability/interpretability, and fostering a human-machine partnership. Finally, we survey some of the more notable applications of DL to tasks across the spectrum of CVI modalities and discuss the future of DL for cardiovascular imaging.

Conclusion and Relevance:

DL is poised to make a meaningful impact on the field of CVI. Rather than a threat, DL should be seen as a partner to the cardiovascular imager in improving efficiency and quality of care.

Introduction:

Automation has played a role in cardiovascular imaging (CVI) for decades, but we are entering a new era of computer systems powered by artificial intelligence (AI) poised to transform the field as we know it. AI systems are computer systems designed to perform tasks that normally require human intelligence. AI systems historically utilized rules-based algorithms to accomplish a specific task, known as symbolic or classical AI. However, the recent popularity of AI has been propelled by the development of *machine learning* technologies, particularly *deep learning.* Machine learning (ML) involves computer algorithms designed to learn without explicit programming, instead relying on patterns and inference from data. Deep learning (DL) is a subset of ML that utilizes a specific type of model called artificial neural networks (ANN).[1] The relationship between AI, ML, and DL is depicted in Supplemental Figure 1.

Deep learning has been revolutionary for many industries, instrumental in innovations like self-driving cars and voice-controlled “smart” assistants. DL has also been successfully applied to problems across several medical disciplines, particularly specialities with plentiful image information such as radiology, dermatology, and pathology.[2–4] We are now seeing the emergence of DL technologies in cardiovascular medicine, most prominently in the field of CVI. DL systems have been employed to automate preprocessing or acquisition of images and assist in the interpretation or

analysis of multiple imaging modalities including echocardiography, cardiac magnetic resonance imaging (cMRI), cardiac computed tomography (CCT), cardiac nuclear imaging, and others. Although not traditionally thought of as an image, electrocardiogram waveform data has also had significant advances with the adoption of deep learning. A remarkable feature of DL models is that they can detect subtle patterns in complex, unstructured data like raw images that may not be readily apparent to even expert interpreters – all with very little human input in the training process. Echocardiograms generally aren't ordered to detect a patient's age or gender, but a recent paper applying DL to echo could predict age with r = 0.68 and gender with AUC = 0.88[5]. If there are subtle clues to these (admittedly non-clinical) findings, perhaps there are other features, obscure to the interpreter, that can provide important diagnostic and prognostic information.

As DL systems become more ubiquitous in CVI, it is imperative that clinicians have a general understanding and familiarity with the underlying methodology. Our goal is to explain the concepts central to understanding how DL models are deployed to accomplish specific tasks in CVI in a simple and digestible fashion towards demystifying these transformative technologies. For the interested reader, a more detailed review of the history of DL and methodological considerations is included in the accompanying supplemental manuscript.

Historical Primer on Artificial Intelligence and Deep Learning:

Although AI systems have recently garnered significant attention in popular media, AI is not a new field. The term "artificial intelligence" was first coined by John McCarthy in 1955. Since that time, the field of AI has undergone a significant evolution with several cycles of considerable hype and industry funding, or 'AI summers', punctuated by periods of low interest and underfunding, or 'AI winters' (Figure 1). It is useful to consider the current explosion in enthusiasm for AI in the context of this volatile history. The current DL boom has been buoyed by three chief technical breakthroughs – major algorithmic advances, the digital "big data" revolution, and significant growth in computational power.

In 2012, the AlexNet model by Alex Krizhevsky and Geoffrey Hinton was a result of these three forces colliding in the use of novel technical approaches (a deep neural network) trained on a large-scale, public image database of over 14 million annotated images called ImageNet using modern, parallel processors called graphics processing units (GPUs).[6,7] Complex neural networks have been previously proposed but were difficult to train without the confluence of large datasets and increasing availability of computational power. The AlexNet model significantly outperformed the previous state-of-the-art benchmark for image classification, cutting the error rate by 50%. This was a watershed moment, as the birth of the modern age of deep learning for computer vision and many of the concepts applying deep learning to CVI that we will discuss in this review stemmed from this effort (Figure 1B).

How Does a Machine Learn?

"Machine learning" evokes images of autonomous robots leading to the downfall of humanity as portrayed in popular media. This has likely contributed to the hesitance and unease on the part of many clinicians in accepting modern AI technologies – a machine that *learns* can be seen as a potential replacement for clinical reasoning and a threat to the clinical workforce itself.[8] This misguided conception is founded in a lack of understanding of ML. Indeed, 'under the hood', a machine *learns* simply by solving a number (actually billions or more) of simple math problems specific to a defined task.

Table 1 provides some definitions of terminology used in ML. There are two general categories of ML algorithms: *supervised* and *unsupervised*. *Supervised* learning involves learning a function that maps a specific input to an output using "labeled" or annotated training data. This form of learning mimics the way a child learns to identify a cat or a dog, by being shown many examples over time and learning from mistakes. This is the more common form of ML in CVI and is typically used for classification, regression, or segmentation tasks – for example to segment the left ventricular cavity on cardiac MRI using expert contours as labels. *Unsupervised* learning involves learning a function via inferences on input data without specific labels. This type of ML can be used to cluster data based on similarities of features (cluster analysis), summarize features of data into a smaller number of learned features (dimensionality reduction), or to generate new data based on features of the input data (generative learning). For example, this type of ML model could be used to cluster visually similar clips from echocardiography to identify different echo views. Though these categories are not mutually exclusive (semi-supervised learning uses partially labeled data), and other forms of ML exist (e.g. reinforcement learning), it can be helpful to conceptualize ML algorithms in this framework.

In the context of computer vision for CVI, the *features* used to train the model are typically numerical representations of the actual pixels from images (Supplemental Figure 2). *Learning* proceeds via a mathematical algorithm that adjusts *parameters* (akin to regression coefficients in statistics) in order to map the numerical representations of input images to the output labels by minimizing the *loss function* – a mathematical representation of the overall error of the model prediction.

What is a Deep Neural Network?

The "deep" in DL refers to the fact that these models rely on ANNs, which are systems of interconnected *neurons* (a reference to the structure being similar to the human nervous system) stacked "deeply" in multiple layers to form increasingly meaningful representations of data as it is transformed by each successive layer. DL models can handle input data with complex features, such as raw images or free text without the need for significant preprocessing. This contrasts with the other "shallow" machine learning models, which often require manual preprocessing of complex data (e.g.

calculating histogram of pixel intensities from an image) to form meaningful representations, a step called *feature engineering.*

The ANNs used in DL consist of *neurons* or *nodes* representing individual features connected in a web like structure via parameters which are called *weights* (Figure 2). DL models that consist solely of these types of simple ANN layers are called fully connected neural networks. However, to handle multidimensional imaging data, more sophisticated DL models are required.

Convolutional neural networks (CNNs) are the basic building blocks of most modern DL computer vision systems. CNNs are well-suited for imaging as they can extract important spatially invariant features from image data. CNNs use a series of filters (or kernels) that slide across the image to detect features relevant to a certain task (such as an edge at a certain orientation, Figure 3). CNNs often include fully connected layers prior to the final output for classification or regression. Fully convolutional neural networks (FCNNs) are special types of CNNs that consist solely of convolutional layers throughout the entire network and are typically used for segmentation tasks as they can reproduce a masked form of the original image by classifying each individual pixel of an image – a process known as semantic segmentation. Other DL architectures that have been deployed to analyze CVI data are reviewed in Supplemental Table 1.

Training and Evaluating a Deep Learning Model

DL models often contain millions of parameters/weights – the loss functions of these models are therefore unknown complex multidimensional ("million-dimensional") surfaces. Consequently, training a DL model requires a mathematical algorithm that performs local optimization to minimize the loss/error - this is typically performed via an iterative algorithm called *gradient descent* (Figure 4). To conceptualize gradient descent, one can imagine a hiker on a mountaintop with heavy cloud cover so that the hiker can only see a few feet in front of him/her. If the goal of the hiker is to descend to the base of the mountain as quickly as possible, he/she will take a step in the direction towards the steepest downward slope. In the same way, the machine calculates the slope of the loss function at the current state of the model, then adjusts the parameters of the model to take a "step" down this slope. Training of the model involves iteratively exposing the model to new examples, with the machine adjusting the model parameters via gradient descent to build a better performing model over time.

In addition to the parameters of the model, there are design choices that must be made when developing any DL model, such as the model size and architecture, the implementation of mathematical optimization (e.g. step size in gradient descent, a.k.a. the *learning rate*), and the number of examples to expose the model to at each iteration (*batch size*). These are termed *hyperparameters* as they are explicitly set by the ML engineer rather than learned by the model. By tuning hyperparameters, a model could theoretically be built that fits inputs of the data to outputs perfectly; however, such a model would be unlikely to work well with new, unseen data. This phenomenon is

called *overfitting*. Overfitting can be thought of as memorizing the input data rather than learning meaningful representations of the data. The same way a student who has memorized the answer key to a test rather than understanding the underlying concepts of the material would perform poorly on new, unseen questions, a model that has overfit to input data would perform poorly on data that it has not been exposed to during training.

To assess for overfitting during training, models must be evaluated in a *validation* or *development* dataset – either by setting aside a proportion of the data or by iteratively partitioning training data ("k-fold cross validation"). Importantly, the "validation" dataset in DL is NOT equivalent to "validation" cohorts used in the statistical risk prediction literature. As hyperparameters are adjusted during training, information from the validation dataset can leak into this process and one can begin to overfit to the validation data itself. Thus, it is crucial to assess model performance of any DL system designed for clinical use on an additional final hold-out *test set* (ideally from an external source) that the model was never exposed to during training.

Challenges and Limitations of Deep Learning for Cardiovascular Imaging

For a DL system to be suitable for widespread clinical implementation, it must be highly accurate, generalizable, unbiased, and interpretable. There are several important barriers to achieving these goals using DL technologies for CVI that deserve mention.

First, although there is active research into developing DL models that can learn from fewer examples ("few shot" learners) or from unlabeled/partially labeled data, DL algorithms traditionally required large, labeled datasets to produce accurate results. Curating large CVI datasets is limited by privacy concerns in ensuring data is deidentified and the logistics of acquiring Digital Imaging and Communications in Medicine (DICOM) images from picture archiving and communication systems (PACS) or vendor neutral archives (VNA). Additionally, labeling data often requires a significant time commitment from domain experts or creative methods to produce labels in an automated fashion, for example using natural language processing of imaging reports. There is no ImageNet for medicine; however, we are starting to see the emergence of large publicly available labeled datasets such as the chest X-ray images provided by the National Institutes of Health,[9] Stanford University,[10] and the Massachusetts Institute of Technology[11] now totaling nearly 1 million images. In CVI, the recently released EchoNet-Dynamic dataset provides over 10,000 apical 4-chamber view clips from echocardiograms performed at Stanford University[12] and the UK Biobank has set an ambitious goal to make 100,000 cardiac MRIs from volunteers aged 40-69 available by the year 2022.[13] As the trend of large publicly available datasets in CVI continues, we will continue to see the growth of more accurate and more clinically useful DL systems.

Even when trained on larger datasets, DL models, due to their complexity, are prone to overfitting which can limit generalizability to external datasets. This opens the door to the dangerous possibility that a model could fit to features irrelevant to the intended task, especially if one is not careful about data curation, which in the context of CVI can have life-threatening consequences. Fortunately, there are a number of strategies to prevent overfitting of DL models such as *transfer learning*, *data augmentation*, *regularization*, *dropout*, and *model ensembling* (Table 1, Supplemental Figure 6) that are discussed in more detail in the supplemental materials.

Moreover, DL models are often referred to as a "black box" – given their complexity it can be difficult to ascertain how a model arrived at a particular prediction. Clinicians are less likely to accept a prediction provided by a ML algorithm if they do not understand the reasoning behind this prediction.[14] Even more concerning is the possibility of introducing human implicit bias when developing a DL model – there are a number of cautionary examples of ML models which resulted in systemic discrimination.[15–19] Therefore, investigators must remain vigilant in ensuring that any algorithm leads to equitable and unprejudiced predictions. Thankfully, research is active into interpretable and explainable DL methods for computer vision, including a collection of visualization methods called saliency mapping, which produce a heatmap highlighting the most important features in an image for the desired output (e.g. the popular gradCAM method[20]).

Even a model that is unbiased and highly accurate at one institution may perform poorly at an external institution due to differences in patient population, imaging equipment (different vendors or machines), or acquisition protocols, a phenomenon known as *dataset shift*.[21] Creative solutions for dataset shift are emerging, including federated learning by sharing DL models across institutions rather than the datasets themselves[22] and generative modeling for image to image translation from one vendor to another.[23] There is, however, no substitute for repeated validation of algorithms in external, never before-seen datasets. For an algorithm to be accepted for clinical use, it must be prospectively validated at multiple institutions using imaging data from multiple vendors to ensure the robustness of its predictions.

Finally, the human-machine interface is crucial to the success of DL technologies for CVI. For a DL system to be successful, it must be adopted by its intended users. This means that it should assimilate nicely into the cardiac imager's workflow and there should be some level of trust and understanding in the decision-making process of the DL system– indeed this is in part the intended goal of this review. Popular culture has promoted a general fear that DL will lead to automation and the replacement of the human workforce, which may lead to hesitance among clinicians in adopting the technology. However, this is generally unfounded as most artificially intelligent systems are designed to *augment* rather than replace human behavior and/or decision making. This concept of "augmented intelligence" is supported by numerous examples in the medical literature - a human being with the

assistance of an artificially intelligent system almost always outperforms the decision-making of a human or machine alone.[24]

Examples of Deep Learning Applications in Cardiovascular Imaging

The following is a brief survey of a selection of the more notable efforts in DL for different modalities in CVI (Table 2). This is not meant to be an exhaustive or systematic review, but rather to highlight some practical examples of the concepts previously discussed.

Echocardiography

Echocardiography presents a unique challenge for DL algorithms - the majority of information is encoded in video clips and different views (i.e. parasternal long axis, apical 4-chamber, etc.) contained in a study are not explicitly labeled as in other imaging modalities. Moreover, though ultrasound can provide better temporal and spatial resolution than CT or MRI, echocardiographic images have a lower signal-to-noise ratio than these modalities and can exhibit significant variation in quality depending on acoustic windows.

Prior efforts have therefore required a comprehensive “bottom up” approach to DL for echocardiographic images with the initial task being to identify the echo view.[5,25,26] Others have focused on view quality assessment by training a CNN on a reference standard from annotations by expert cardiologists.[26,27]

Most applications of DL to echocardiographic images to date have focused on cardiac segmentation or assessment of ventricular structure and function rather than identifying pathology or predicting risk, though there are a few exceptions. Ghorbani et al.[5] used individual frames to identify images with cardiac pathology (e.g. left atrial dilation, pacemaker lead) in addition to patient demographics (e.g. age, weight, sex) albeit with variable performance (AUC 0.75-0.89, R^2 0.33-0.73). This is an important reminder of the limitations of DL – if a patient’s age cannot be definitively determined by echocardiography, a DL algorithm is not likely to converge on a solution. Zhang et al.[26] presented a fully automated pipeline for the processing and interpretation of echocardiograms using over 14,000 studies, using a series of CNNs and FCNNs to perform segmentation of cardiac structures, calculate chamber volumes and functional information, and distinguish hypertrophic cardiomyopathy, cardiac amyloidosis, and pulmonary arterial hypertension from normal controls. Others have presented DL models for identifying left ventricular hypertrophy or cardiac amyloidosis in a high throughput fashion on large multi-institutional datasets.[28–30]

While most DL applications in echocardiography treat individual frames of an echo clip as data inputs, Ouyang et al.[31] introduced the first truly “video-based” DL algorithm for assessment of left

ventricular ejection fraction using a 3-dimensional CNN with spatio-temporal convolutions that outperformed previously published results. Notably, the authors utilized transfer learning by pretraining their 3D CNN on a publicly available human action video dataset.

Although DL for echocardiography is still in its nascence, there are already several commercial products in this space. For example, Caption Health's innovative EchoGPS platform uses a DL algorithm to guide the *acquisition* of images rather than the interpretation with the goal of allowing less experienced operators acquire diagnostic quality images.[32] However, there remain many unexplored applications - to date, very few investigators have applied DL methodology to the analysis of 3-dimensional echocardiography or Doppler echocardiography.

Cardiac MRI and Cardiac CT

Cardiac MRI and CCT present unique challenges for DL analysis as acquisitions are by nature 3-dimensional volumes rather than 2D-images. While many investigators have utilized 2D data (pixels) on a slice-by-slice basis to train their DL algorithms, often superior results can be achieved by accounting for the 3-dimensional relationship of the data (voxels). Additionally, cMRI and CCT data is captured with a high dynamic range (a large range of gray intensities) requiring additional preprocessing of data prior to analysis.

There has been more work published related to DL analysis of cMRI than any other CVI modality, but, the majority have focused on cardiac chamber segmentation of cine loops for structural and functional evaluation. This is likely due to the combination of a number of high quality publicly available cMRI datasets as well as online data science competitions which have attracted popular interest in tackling this problem. One of the largest segmentation efforts to date utilized data from almost 5,000 cMRIs from the UK Biobank to train a FCNN to segment left/right ventricular and atrial cavities with impressive accuracy (<10mL mean absolute difference between automated and manual volume estimation for all chambers).[33] Circle CVI42's commercial platform has integrated a similar algorithm for automated cardiac chamber segmentation.

Others have focused on segmenting different MRI sequences or have targeted tasks other than segmentation. Fahmy et al.[34] applied a FCNN to the task of segmenting T1 sequences towards fully automated T1 mapping. Interestingly, Zhang et al.[35] developed an ambitious algorithm with the goal of going beyond what's possible with the human eye – myocardial scar delineation based on regional wall motion patterns *without* the use of gadolinium contrast using a model that combined multiple DL architectures to translate local and global motion features into a pixelwise prediction of myocardial scarring.

For CCT, most investigations to date have focused on automated coronary CT angiography (CCTA) analysis or calcium scoring. Coennen et al.[36] tested a DL-based algorithm designed to estimate non-invasive CCTA-based fractional flow reserve (FFR) in a population of 351 patients which dramatically improved assessment of functionally obstructive coronary artery disease when compared with visual assessment (58% vs. 78% accuracy on a per-vessel basis). Eisenberg et al.[37] innovatively developed one of the first image-based clinical risk prediction DL models to date by employing a previously established FCNN for intrapericardial segmentation on CCT to form an attenuation map for identifying epicardial adipose tissue volume, which was additive to coronary calcium scoring in predicting major adverse cardiovascular events.

Perhaps the most innovative uses for DL for cMRI and CCT have been generative learning applications for reconstruction, artifact attenuation, or denoising of images. Wolterink et al.[38] utilized a generative adversarial network (GAN) to denoise low-dose CT images by first training on phantom image data then fine tuning on clinically acquired CTs for coronary calcium scoring, producing results similar to full-dose CT acquisitions. Fan et al.[39] developed a FCNN for reconstruction of undersampled cardiac perfusion MRI k-space data that reduced reconstruction time up to 14.4-fold compared to the more computationally expensive compressed sensing method. This highlights an important principle of DL – while training a DL model is computationally demanding, *inference* using a trained model is extremely rapid compared to traditional rules-based algorithms.

Cardiac Nuclear Imaging and Other Imaging Modalities

Nuclear imaging data is relatively lower resolution than the other imaging modalities discussed here and thus may not require an additional preprocessing step to downsample images to a lower resolution before feeding the data into a DL algorithm as is the case with echo, cMRI, and CCT. However, nuclear imaging traditionally requires significant preprocessing for clinical interpretability - for example the generation of perfusion polar maps in myocardial perfusion imaging. Betancur et al.[40,41] developed a CNN to determine per-patient and per-vessel ischemia from raw, black-out, and quantitative polar maps. When compared to standard total perfusion deficit calculated by comparison to normal reference values, the DL model resulted in modest improvement (per patient AUC 0.80 vs. 0.78). This is an important observation – image preprocessing to produce polar maps can be thought of as manual feature engineering and results in a relatively limited feature space (64x64 pixels); thus DL likely provides limited benefit over simpler modes of analysis using rules-based methods. Instead, leveraging the ability of DL models to automatically extract important features from raw single photon emission computed tomography (SPECT) data, for example inferring areas of attenuation artifact, [42] might lead to superior performance. This concept was demonstrated by Jun et al.[43] in their analysis of coronary intravascular ultrasound. A CNN trained to predict areas of thin capped fibroatheroma (using optical coherence tomography as a gold standard) on raw images significantly

outperformed “shallow” ML methods that required manual feature extraction as an initial preprocessing step (AUC 0.911 vs. 0.844-0.859). For coronary angiography, Avram et al. recently presented a first of its kind DL system consisting of a series of CNNs for the fully automated detection of the projection angle, recognition of coronary artery anatomy, and identification of obstructive coronary lesions.[44]

Future of Deep Learning for Cardiovascular Imaging:

Relative to other industries, DL for medical imaging and particularly for CVI is still in its infancy. The majority of published results thus far have been on retrospective datasets. The future of DL for CVI will include prospective validation of DL systems in well-designed randomized clinical trials. Only then will we begin to see actual widespread clinical adoption of these technologies.

In addition, the paradigm of using DL to automate tasks that can easily be accurately performed by a human operator will likely begin to shift to tackling problems that have thus far evaded a solution, such as image-based risk prediction. We have only recently seen the emergence of generative learning for image reconstruction, denoising, artifact attenuation, and super-resolution – we are likely to see more advanced methodologies and more creative uses of this powerful technology in the coming years. We will also likely see more DL systems that integrate multimodal imaging and non-imaging data.

One area that has thus far received relatively little attention in CVI is natural language processing (NLP). While images contain important features for disease identification and prognosis prediction, the communication of those features takes place through natural language. NLP can be used to extract labels from imaging reports for training computer vision algorithms or conversely to caption images for automated report generation. NLP can also be used as a tool for a clinician (or even a patient) to interface with imaging data, akin to the popular “smart-assistants” in many households. While most NLP research related to medical imaging has historically used rules-based methods with regular expressions, the field of NLP has been revolutionized with the introduction of deep self-attention models called transformers[45] (e.g. BERT,[46] GPT-3[47]) which have smashed previously set benchmarks for accuracy on common NLP tasks. Deep learning-based NLP is likely to become a target of intense research in CVI.

Finally, while we have observed here that DL algorithms can simulate remarkable intuition or “gestalt” in identifying anatomic or structural correlates, we should not forget that the world is in the end governed by definable physical principles. While DL models do not require preprocessing of features, incorporating domain knowledge regarding these underlying physical principles can lead to boosts in performance, efficiency, and generalizability of DL algorithms for CVI analysis.[48]

Moreover, not every problem requires a DL solution – often the more accurate and more efficient approach is the simpler one. After all, the Bernoulli equation unequivocally connects aortic valve velocity to gradient via a quadratic relationship – it is not necessary to use a million-parameter model to derive this formula via gradient descent. The future of AI for CVI is likely to be dominated by *hybrid* neuro-symbolic algorithms[49] that incorporate DL technologies with classical rules-based programming while accounting for underlying physical principles of imaging and physiology.

Conclusion:

Deep learning is poised to make a meaningful impact on the field of cardiovascular imaging. Though the field is still blossoming, we have already seen remarkable applications of DL towards improving efficiency and accuracy of image interpretation, acquisition, and processing. To adapt a quote by Curtis Langlotz,[50] deep learning will not replace cardiovascular imagers, but cardiovascular imagers who use and understand deep learning technologies will replace those who do not.

References

1. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.

2. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, Lungren MP, Ng AY. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *http://arxiv.org/abs/171105225*. 2017;

3. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115–118.

4. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, Brogi E, Reuter VE, Klimstra DS, Fuchs TJ. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*. 2019;25:1301–1309.

5. Ghorbani A, Ouyang D, Abid A, He B, Chen JH, Harrington RA, Liang DH, Ashley EA, Zou JY. Deep learning interpretation of echocardiograms. *npj Digital Medicine*. 2020;3:1–10.

6. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks.

7. Deng J, Dong W, Socher R, Li L-J, Kai Li, Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition. Institute of Electrical and Electronics Engineers (IEEE); 2009. p. 248–255.

8. Sarwar S, Dent A, Faust K, Richer M, Djuric U, van Ommeren R, Diamandis P. Physician perspectives on integration of artificial intelligence into diagnostic pathology. *npj Digital Medicine*. 2019;2:1–7.

9. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. 2017. p. 3462–3471.

10. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, Marklund H, Haghgoo B, Ball R, Shpanskaya K, Seekins J, Mong DA, Halabi SS, Sandberg JK, Jones R, Larson DB, Langlotz CP, Patel BN, Lungren MP, Ng AY. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison [Internet].

11. Johnson AEW, Pollard TJ, Berkowitz SJ, Greenbaum NR, Lungren MP, Deng C, Mark RG, Horng S. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific Data 2019 6:1*. 2019;6:1–8.

12. Ouyang D, Ashley EA. EchoNet-Dynamic : a Large New Cardiac Motion Video Data Resource for Medical Machine Learning. In: NeurIPS ML4H Workshop. 2019. p. 1–11.

13. Littlejohns TJ, Holliday J, Gibson LM, Garratt S, Oesingmann N, Alfaro-Almagro F, Bell JD, Boultwood C, Collins R, Conroy MC, Crabtree N, Doherty N, Frangi AF, Harvey NC, Leeson P, Miller KL, Neubauer S, Petersen SE, Sellors J, Sheard S, Smith SM, Sudlow CLM, Matthews PM, Allen NE. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nature Communications 2020 11:1*. 2020;11:1–12.

14. Diprose WK, Buist N, Hua N, Thurier Q, Shand G, Robinson R. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. *Journal of the American Medical Informatics Association*. 2020;27:592–600.

15. Gianfrancesco MA. Potential Biases in ML Algorithms Using EHR Data. 2019;178:1544–1547.

16. Hajian S, Bonchi F, Castillo C. Algorithmic bias: From discrimination discovery to fairness-aware data mining. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016;13-17-Augu:2125–2126.

17. Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proceedings of the National Academy of Sciences of the United States of America*. 2020;117:12592–12594.

18. Agüera B, Hee A, Ryu J, Adam H, Lovejoy J, Hovy D, Benton A, Lemoine B, Zhang B, Beutel A, Zaldivar A, Wu S, Barnes P, Vasserman L, Hutchinson B, Spitzer E, Raji D. Bias in the Vision and Language of Artificial Intelligence.

19. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447–453.

20. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision*. 2017;2017-Octob:618–626.

21. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, Kohane IS, Saria S. The Clinician and Dataset Shift in Artificial Intelligence. *New England Journal of Medicine*. 2021;385:283–286.

22. Sarma K v., Harmon S, Sanford T, Roth HR, Xu Z, Tetreault J, Xu D, Flores MG, Raman AG, Kulkarni R, Wood BJ, Choyke PL, Priester AM, Marks LS, Raman SS, Enzmann D, Turkbey B, Speier W, Arnold CW. Federated learning improves site performance in multicenter deep learning without data sharing. *Journal of the American Medical Informatics Association*. 2021;28:1259–1264.

23. Marcadent S, Hofmeister J, Preti MG, Martin SP, van de Ville D, Montet X. Generative Adversarial Networks Improve the Reproducibility and Discriminative Power of Radiomic Features. *Radiology: Artificial Intelligence*. 2020;2:e190035.

24. Patel BN, Rosenberg L, Willcox G, Baltaxe D, Lyons M, Irvin J, Rajpurkar P, Amrhein T, Gupta R, Halabi S, Langlotz C, Lo E, Mammarappallil J, Mariano AJ, Riley G, Seekins J, Shen L, Zucker E, Lungren MP. Human–machine partnership with artificial intelligence for chest radiograph diagnosis. *npj Digital Medicine*. 2019;2.

25. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. *npj Digital Medicine*. 2018;1:6.

26. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan CR, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo RC. Fully automated echocardiogram interpretation in clinical practice: Feasibility and diagnostic accuracy. *Circulation*. 2018;138:1623–1635.

27. Abdi AH, Luong C, Tsang T, Allan G, Nouranian S, Jue J, Hawley D, Fleming S, Gin K, Swift J, Rohling R, Abolmaesumi P. Automatic Quality Assessment of Echocardiograms Using Convolutional Neural Networks: Feasibility on the Apical Four-Chamber View. *IEEE Transactions on Medical Imaging*. 2017;36:1221–1230.

28. Duffy G, Cheng PP, Yuan N, He B, Kwan AC, Shun-Shin MJ, Alexander KM, Ebinger J, Lungren MP, Rader F, Liang DH, Schnittger I, Ashley EA, Zou JY, Patel J, Witteles R, Cheng S, Ouyang D. High-Throughput Precision Phenotyping of Left Ventricular Hypertrophy with Cardiovascular Deep Learning.

29. Howard JP, Stowell CC, Cole GD, Ananthan K, Demetrescu CD, Pearce K, Rajani R, Sehmi J, Vimalesvaran K, Kanaganayagam GS, McPhail E, Ghosh AK, Chambers JB, Singh AP, Zolgharni M, Rana B, Francis DP, Shun-Shin MJ. Automated Left Ventricular Dimension Assessment Using Artificial Intelligence Developed and Validated by a UK-Wide Collaborative. *Circulation: Cardiovascular Imaging*. 2021;405–415.

30. Goto S, Mahara K, Beussink-Nelson L, Ikura H, Katsumata Y, Endo J, Gaggin HK, Shah SJ, Itabashi Y, MacRae CA, Deo RC. Artificial intelligence-enabled fully automated detection of cardiac amyloidosis using electrocardiograms and echocardiograms. *Nature Communications*. 2021;12.

31. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA, Harrington RA, Liang DH, Ashley EA, Zou JY. Video-based AI for beat-to-beat assessment of cardiac function. *Nature*. 2020;580.

32. Narang A, Bae R, Hong H, Thomas Y, Surette S, Cadieu C, Chaudhry A, Martin RP, McCarthy PM, Rubenson DS, Goldstein S, Little SH, Lang RM, Weissman NJ, Thomas JD. Utility of a Deep-Learning Algorithm to Guide Novices to Acquire Echocardiograms for Limited Diagnostic Use. *JAMA Cardiology*. 2021;6:624–632.

33. Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, Lee AM, Aung N, Lukaschuk E, Sanghvi MM, Zemrak F, Fung K, Paiva JM, Carapella V, Kim YJ, Suzuki H, Kainz B, Matthews PM, Petersen SE, Piechnik SK, Neubauer S, Glocker B, Rueckert D. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*. 2018;20.

34. Fahmy AS, El-Rewaidy H, Nezafat M, Nakamori S, Nezafat R. Automated analysis of cardiovascular magnetic resonance myocardial native T1 mapping images using fully convolutional neural networks. *Journal of Cardiovascular Magnetic Resonance*. 2019;21.

35. Zhang N, Yang G, Gao Z, Xu C, Zhang Y, Shi R, Keegan J, Xu L, Zhang H, Fan Z, Firmin D. Deep learning for diagnosis of chronic myocardial infarction on nonenhanced cardiac cine MRI. *Radiology*. 2019;291:606–607.

36. Coenen A, Kim YH, Kruk M, Tesche C, de Geer J, Kurata A, Lubbers ML, Daemen J, Itu L, Rapaka S, Sharma P, Schwemmer C, Persson A, Schoepf UJ, Kepka C, Yang DH, Nieman K. Diagnostic accuracy of a machine-learning approach to coronary computed tomographic

angiography–Based fractional flow reserve result from the MACHINE Consortium. *Circulation: Cardiovascular Imaging*. 2018;11:1–11.

37. Eisenberg E, McElhinney PA, Commandeur F, Chen X, Cadet S, Goeller M, Razipour A, Gransar H, Cantu S, Miller RJH, Slomka PJ, Wong ND, Rozanski A, Achenbach S, Tamarappoo BK, Berman DS, Dey D. Deep Learning-Based Quantification of Epicardial Adipose Tissue Volume and Attenuation Predicts Major Adverse Cardiovascular Events in Asymptomatic Subjects. *Circulation Cardiovascular imaging*. 2020;13:e009829.

38. Wolterink JM, Leiner T, Viergever MA, Išgum I. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Transactions on Medical Imaging*. 2017;36:2536–2545.

39. Fan L, Shen D, Haji-Valizadeh H, Naresh NK, Carr JC, Freed BH, Lee DC, Kim D. Rapid dealiasing of undersampled, non-Cartesian cardiac perfusion images using U-net. *NMR in Biomedicine*. 2020;33:1–8.

40. Betancur J, Commandeur F, Motlagh M, Sharir T, Einstein AJ, Bokhari S, Fish MB, Ruddy TD, Kaufmann P, Sinusas AJ, Miller EJ, Bateman TM, Dorbala S, di Carli M, Germano G, Otaki Y, Tamarappoo BK, Dey D, Berman DS, Slomka PJ. Deep Learning for Prediction of Obstructive Disease From Fast Myocardial Perfusion SPECT: A Multicenter Study. *JACC: Cardiovascular Imaging*. 2018;11:1654–1663.

41. Hu L-H, Betancur J, Sharir T, Einstein AJ, Bokhari S, Fish MB, Ruddy TD, Kaufmann PA, Sinusas AJ, Miller EJ, Bateman TM, Dorbala S, di Carli M, Germano G, Commandeur F, Liang JX, Otaki Y, Tamarappoo BK, Dey D, Berman DS, Slomka PJ. Machine learning predicts per-vessel early coronary revascularization after fast myocardial perfusion SPECT: results from multicentre REFINE SPECT registry. *European Heart Journal - Cardiovascular Imaging*. 2019;

42. Yang J, Shi L, Wang R, Miller EJ, Sinusas AJ, Liu C, Gullberg GT, Seo Y. Direct Attenuation Correction Using Deep Learning for Cardiac SPECT: A Feasibility Study. *Journal of Nuclear Medicine*. 2021;62:1645–1652.

43. Jun TJ, Kang SJ, Lee JG, Kweon J, Na W, Kang D, Kim D, Kim D, Kim YH. Automated detection of vulnerable plaque in intravascular ultrasound images. *Medical and Biological Engineering and Computing*. 2019;57:863–876.

44. Avram R, Olgin JE, Wan A, Ahmed Z, Verreault-Julien L, Abreau S, Wan D, Gonzalez JE, So DY, Soni K, Tison GH. CathAI: Fully Automated Interpretation of Coronary Angiograms Using Neural Networks. 2021;

45. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. *Advances in Neural Information Processing Systems*. 2017;2017-Decem: 5999–6009.

46. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2018;

47. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler DM, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, Mccandlish S, Radford A, Sutskever I, Openai DA. Language Models are Few-Shot Learners. 2020.

48. Litjens G, Ciompi F, Wolterink JM, de Vos BD, Leiner T, Teuwen J, Išgum I. State-of-the-Art Deep Learning in Cardiovascular Image Analysis. *JACC: Cardiovascular Imaging*. 2019;12:1549–1565.

49. Mao J, Gan C, Kohli P, Tenenbaum JB, Wu J. The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision. *7th International Conference on Learning Representations, ICLR 2019*. 2019;

50. Reardon S. Rise of Robot Radiologists. *Nature*. 2019;576:S54–S58.

51. Moradi M, Guo Y, Gur Y, Negahdar M, Syeda-Mahmood T. A Cross-Modality Neural Network Transform for Semi-automatic Medical Image Annotation. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, editors. Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2016. Cham: Springer International Publishing; 2016. p. 300–307.

52. Diller GP, Babu-Narayan S, Li W, Radojevic J, Kempny A, Uebing A, Dimopoulos K, Baumgartner H, Gatzoulis MA, Orwat S. Utility of machine learning algorithms in assessing patients with a systemic right ventricle. *European Heart Journal Cardiovascular Imaging*. 2019;20:925–931.

53. Tao Q, Yan W, Wang Y, Paiman EHM, Shamonin DP, Garg P, Plein S, Huang L, Xia L, Sramko M, Tintera J, de Roos A, Lamb HJ, van der Geest RJ. Deep learning-based method for fully automatic quantification of left ventricle function from cine MR images: A multivendor, multicenter study. *Radiology*. 2019;290:81–88.

54. Bratt A, Kim J, Pollie M, Beecy AN, Tehrani NH, Codella N, Perez-Johnston R, Palumbo MC, Alakbarli J, Colizza W, Drexler IR, Azevedo CF, Kim RJ, Devereux RB, Weinsaft JW. Machine learning derived segmentation of phase velocity encoded cardiovascular magnetic resonance for fully automated aortic flow quantification. *Journal of Cardiovascular Magnetic Resonance*. 2019;21:1–11.

55. Masutani EM, Bahrami N, Hsiao A. Deep learning single-frame and multiframe super-resolution for cardiac MRI. *Radiology*. 2020;295:552–561.

56. Itu L, Rapaka S, Passerini T, Georgescu B, Schwemmer C, Schoebinger M, Flohr T, Sharma P, Comaniciu D. A machine-learning approach for computation of fractional flow reserve from coronary computed tomography. *Journal of Applied Physiology*. 2016;121:42–52.

57. van Velzen SGM, Lessmann N, Velthuis BK, Bank IEM, van den Bongard DHJG, Leiner T, de Jong PA, Veldhuis WB, Correa A, Terry JG, Carr JJ, Viergever MA, Verkooijen HM, Išgum I. Deep learning for automatic calcium scoring in CT: Validation using multiple cardiac CT and chest CT protocols. *Radiology*. 2020;295:66–79.

58. Lessmann N, van Ginneken B, Zreik M, de Jong PA, de Vos BD, Viergever MA, Isgum I. Automatic Calcium Scoring in Low-Dose Chest CT Using Deep Neural Networks with Dilated Convolutions. *IEEE Transactions on Medical Imaging*. 2018;37:615–625.

59. Commandeur F, Goeller M, Razipour A, Cadet S, Hell MM, Kwiecinski J, Chen X, Chang H-J, Marwan M, Achenbach S, Berman DS, Slomka PJ, Tamarappoo BK, Dey D. Fully Automated CT Quantification of Epicardial Adipose Tissue by Deep Learning: A Multicenter Study. *Radiology: Artificial Intelligence*. 2019;1:e190045.

60. Mu D, Bai J, Chen W, Yu H, Liang J, Yin K, Li H, Qing Z, He K, Yang H-Y, Zhang J, Yin Y, McLellan HW, Schoepf UJ, Zhang B. Calcium Scoring at Coronary CT Angiography Using Deep Learning. *https://doi.org/101148/radiol2021211483*. 2021;

61. Betancur J, Commandeur F, Motlagh M, Sharir T, Einstein AJ, Bokhari S, Fish MB, Ruddy TD, Kaufmann P, Sinusas AJ, Miller EJ, Bateman TM, Dorbala S, di Carli M, Germano G, Otaki Y, Tamarappoo BK, Dey D, Berman DS, Slomka PJ. Deep Learning for Prediction of Obstructive Disease From Fast Myocardial Perfusion SPECT: A Multicenter Study. *JACC: Cardiovascular Imaging*. 2018;11:1654–1663.

Table 1. Machine/Deep Learning Definitions

Term	Definition
Inputs/examples	The data from the subjects or patients being studied

Output/target	The intended result of the model – e.g. a classification, prediction, reconstructed/enhanced image, or text caption
Features	The predictors or variables derived from each input included in the model – in the context of imaging, the features are often a numerical representation of the pixels of the image
Parameters/ weights	Numerical values that are adjusted during the learning process akin to regression coefficients in statistics.
Loss/cost function	A mathematical function that describes of how far the model output is from the desired output - or the overall error of the model
Learning	A process by which a mathematical optimization algorithm is used to adjust parameters of the model to map features of the inputs to the desired output by minimizing the loss function or error
Labels	The desired or correct output/target for each input of training data
Supervised learning	Learning using “labeled” or annotated training data – e.g. classification, regression, or segmentation tasks
Unsupervised learning	Learning via inferences on input data without explicit labels – e.g. cluster analysis, dimensionality reduction
Gradient Descent	A mathematical optimization algorithm to minimize the error/loss function that calculates the slope of the loss function with respect to each parameter in the model and iteratively takes a “step” down this slope by adjusting parameters.
Hyperparameters	Parameters of the model (e.g. model architecture, learning rate) that are not learned by the model but are instead explicitly set by the machine learning engineer.
Overfitting	The process by which a model fits to noise or irrelevant features in the training dataset rather than robust features that are important for a particular task. Can be thought of as “memorizing the data” rather than learning.

Training Dataset	A dataset on which the model is optimized or from which the model “learns”.
Validation/ Development Dataset	A dataset used to monitor for overfitting during the training process. NOT equivalent to the “validation dataset” in risk prediction literature as it does nothing to validate the model.
Testing Dataset	A dataset that the final, trained model is evaluated on. This should be a dataset that the model is not exposed to at all during the training process, ideally from an external cohort.
Transfer Learning	Pretraining a neural network on a different, but related task or domain, then retraining all or parts of this model for the intended task.
Data Augmentation	Simulating new training examples by applying random transformations to existing data (e.g. random cropping, zooming, stretching, and brightness/contrast shifts to images).
Regularization	A strategy to prevent overfitting that adds a penalty term to the error function tto penalize very large weights. By “shrinking” the weights towards zero, regularization prevents a model from overfitting to noise in the dataset.
Dropout	Randomly dropping (setting to zero) a proportion of neurons in a neural network during each training iteration to prevent overfitting and encourage the model to fit to more robust, generalizable features.
Model Ensembling	Combining the results of multiple weaker performing models into one better performing model.

Table 2. Examples of DL Applications in Cardiovascular Imaging

Modality	Authors	Year	Task	Architecture	Dataset	Performance

Echo	Moradi et al. [51]	2016	Semi-automatic annotation/ report generation of CW-Doppler images	7-layer CNN for image analysis, Fully connected NN for image to text transformation	496 CW Images 10,253 valve descriptions	96% acc for valve 91% acc for regurg
	Zhang et al.[26]	2018	End-to-end - 1) View classification, 2) Segmentation, 3) Structure and Function, 4) Disease Classification	View classification – CNN Segmentation - FCNN Disease classification – CNN	14,035 echos (split for different tasks)	96% acc for view MAD 9.7% for LVEF AUC 0.85-0.93 for disease
	Madani et al.[25]	2019	View classification LVH Classification	CNNs and FCNN GANs for semi-supervised learning	267 echos (220,000 images)	Acc 91.7% for view
	Diller et al.[52]	2019	Identification and segmentation of systemic ventricle in cases of TGA	Identification - Simple 6-layer CNN Segmentation - FCNN	123 patients (>100,000 frames)	Acc 98% for diagnosis Dice seg 0.79-0.88
	Ghorbani et al.[5]	2020	Frame-level view classification. Identification of structure, function, and phenotypes	2D-CNN for image classification	~3,300 echos (>1.6 million images)	AUC 0.75-0.89 R^2 0.33-0.56
	Ouyang et al.[31]	2020	Beat to beat assessment of LVEF on video clips	LVEF Estimation - 3D-CNN (spatiotemporal) + Segmentation - FCNN	>10,000 echo clips	MAE LVEF 4.1% AUC HF 0.97 Dice seg = 0.92
	Howard et al.[29]	2021	Estimate interventricular	CNN for keypoint detection of		

			septal and posterior wall thickness	endocardial and epicardial measurements	2056 individual frames from 1265 clips	MAD 0.9mm for IVd and PWd thickness
	Goto et al.[30]	2021	Fully automated detection of cardiac amyloidosis on echo (and ECG)	3D-CNN	13,177 clips - training and internal testing 1,278 clips from 4 AMCs - external testing	AUC 0.89-1.00
	Narang et al.[32]	2021	Prospective evaluation of DL-based real time guidance for novice operators to obtain diagnostic quality echo images	Series of CNNs for 1)Diagnostic quality, 2) 6D distance to ideal probe position, & 3) Corrective probe manipulations	Trained on >5 million examples; Prospectively valuated on 30 patients scanned by 8 nurses (240 studies)	Diagnostic quality images obtained in 92.5%-98.8% of cases
MRI	Bai et al. 33	2018	Semantic segmentation of LV, RV, LA, and RA	FCNN	4,875 MRIs 93,500 images	Dice seg = 0.88-0.96 MAE - LVEDV 6.1mL, RVEDV 8.5mL
	Tao et al. 53	2019	Segmentation of LV on short-axis cine views	FCNN trained on single-center vs. multi-center multi-vendor datasets	400 patients 41,593 images	Dice seg = 0.88-0.95 (multicenter) vs. 0.82-0.90 (single center)
	Zhang et al.[35]	2019	Delineate areas of scar on non-enhanced MRI images	CNN for image features, LSTM for motion features, stacked autoencoder to classify MI territories	299 patients	Per segment sens/spec for detecting MI = 89.8%/99.1%

	Fahmy et al.[34]	2019	Automated analysis of native T1 mapping	FCNN	210 patients 11,550 images	Dice seg = 0.85 Calculated T1 *R* = 0.82
	Bratt et al.[54]	2019	Automated analysis of phase velocity encoded sequences for aortic flow quantification	FCNN	150 patients 4,345 images Eval 190 patients	*R* = 0.99 for ML and manual derived aortic flow
	Fan et al. 39	2020	Reconstruct de-aliased cardiac perfusion sequences from non-Cartesian k-space	3D FCNN	40 patients >10,000 2D images	SSIM vs. CS = 0.94 14.4 times faster than CS
	Masutani et al.[55]	2020	Super-resolution to enhance detail from faster small-matrix MRI acquisitions	4 different 2D and 3D CNNs/FCNNs for super-resolution	400 MRIs 367 patients	SSIM = 0.76-0.99 CNNs vs. 0.65-0.98 for zero padding
CT	Wolterink et al.[38]	2017	Noise reduction of low dose CCT	GANs	Phantom images 28 patients	CAC scored in 3/28 low-dose images vs. 9/28 without noise reduction
	Coenen et al.[36]	2018	Clinical evaluation of automated non-invasive CCTA-based FFR developed by Itu et al.[56]	Fully connected neural network (6-layers) using features derived from rules-based algorithm as input	351 patients 525 vessels (all for testing)	*R* = 0.997 between ML & CFD-based FFR
	van Velzen et al.[57]	2020	Clinical evaluation of automatic calcium scoring	FCNN followed by CNN	7,240 participants	ICC 0.79-0.97 between automated & manual

			algorithm by Lessman et al.[58]			reference scores for CAC
	Eisenberg et al.[37]	2020	Clinical evaluation of algorithm for CT quantification of epicardial adipose tissue by Commandeur et al. [59]	CNN for slice classification, FCNN for intrapericardial segmentation	2,068 patients	DL-derived EAT volume HR for MACE = 1.35
	Mu et al. [60]	2021	Calcium scoring using contrast-enhanced CCTA	3D-FCNN for calcium segmentation, 3D-CNN for regression to CAC score	Training 292 patients Validation 73 patients Testing 240 CTA scans	R = 0.96 between CAC for DL on CCTA and semi-auto on non-con CT
Others	Betancur et al.[61]	2018	Identify per-patient and per-vessel ischemia from SPECT MPI	CNN using features from polar maps	1,638 patients	Per patient AUC 0.80 vs. 0.78 for standard total perfusion deficit
	Jun et al. [43]	2018	Identify thin-cap fibroatheroma on IVUS (OCT as gold-standard)	CNN compared to manual feature extraction + shallow ML models (KNN, RF)	100 patients 12,325 images	AUC 0.911 for DL vs. 0.844-0.859 for other ML
	Avram et al.[44]	2021	Fully-automated coronary angiogram interpretation to detect obstructive coronary disease	Series of CNNs for 1) detection of the projection angle, 2)recognition of coronary anatomy, and 3)identification of obstructive lesions	1,418,297 images from 11,972 patients and 195,195 videos; Externally tested on 464 angiogram videos	AUC for detection obstructive stenosis (≥ 70%) = 0.869

Abbreviations: CW = continuous wave, CNN = convolutional neural network, NN = neural network, acc = accuracy, regurg = regurgitation, FCNN = fully convolutional neural network, MAD = median absolute deviation, LVEF = left ventricular ejection fraction, AUC = area under the receiver operator characteristic curve, LVH = left ventricular hypertrophy, GAN = generative adversarial network, TGA = transposition of the great arteries, seg = segmentation, MAE = mean absolute error, HF = heart failure, ECG = electrocardiogram, IVd = interventricular thickness in diastole, PWd = posterior wall thickness in diastole, AMC = American medical center, DL = deep learning, LV = left ventricle, RV = right ventricle, LA = left atrium, RA = right atrium, MRI = magnetic resonance imaging, CS = compressed sensing, IVUS = intravascular ultrasound, OCT = optical coherence tomography, ML = machine learning, KNN = k nearest neighbors, RF = random forest, CCTA = coronary computed tomography angiography, CCT = cardiac CT, FFR = fractional flow reserve, SSIM = structural similarity index, MI = myocardial infarction, SPECT = single photon emission computed tomography, MPI = myocardial perfusion imaging, CFD = computational fluid dynamics, *R* = correlation coefficient, CAC = coronary artery calcium, ICC = intraclass correlation coefficient, epicardial adipose tissue, HR = hazard ratio, MACE = major adverse cardiac events

A

B

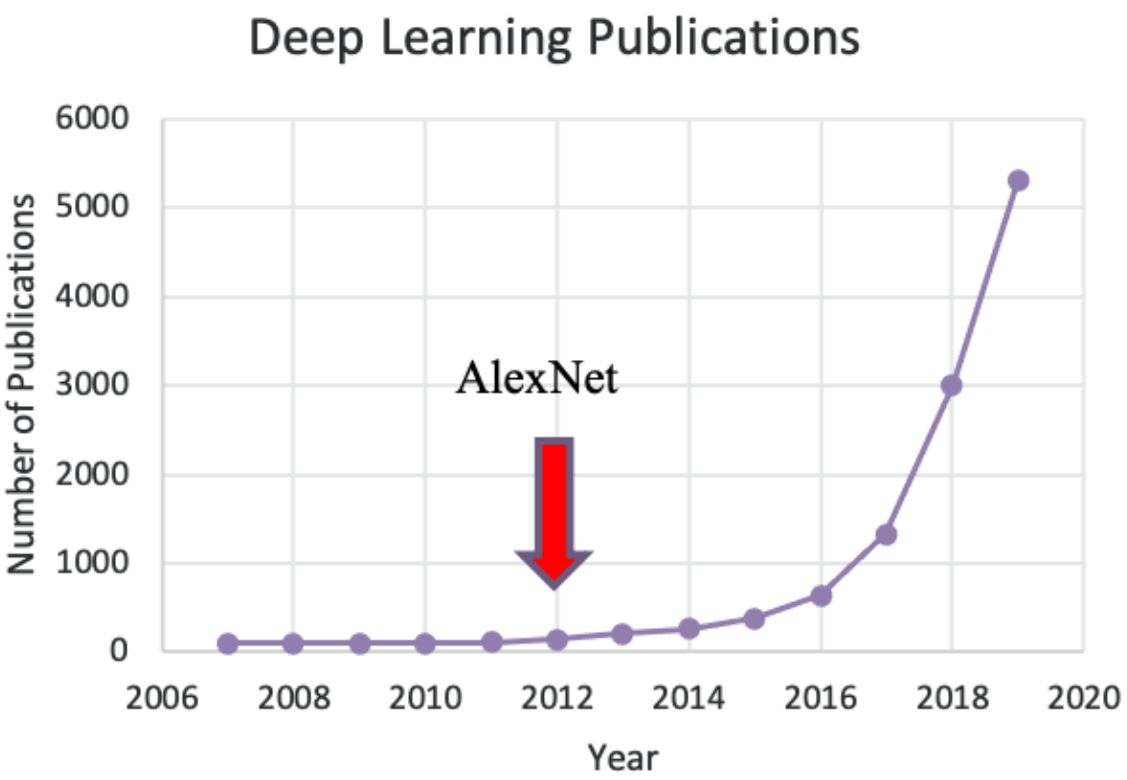

Figure 1. History of AI Through the Lens of Deep Learning

1. A timeline of major events in the history of AI from a DL-centric view. Notably, the field of AI has undergone a significant evolution with several cycles of considerable hype and industry funding, or 'AI summers', punctuated by periods of low interest and underfunding, or 'AI winters'.
2. AlexNet by Alex Krizhesky and Geoffrey Hinton was a major achievement for DL in that it incorporated algorithmic advances allowing scaling and efficiency of a DL model that was trained on a large, labeled dataset (ImageNet) using the parallel processing power of modern graphics processing units (GPUs). This was watershed moment for the field of DL, as evident by the exponential growth in DL publications that followed. Data obtained from *https://pubmed.ncbi.nlm.nih.gov/*.

Figure 2. A representation of an artificial neural network

Simplified schematic of an artificial neural network. The network consists of neurons or nodes arranged in layers and connected by a web of parameters called weights. The input features are transformed in each hidden layer to arrive at the desired output. Each feature node's value is

calculated as the sum of the products of each corresponding input feature and weight. This result is then passed through a non-linear activation function (otherwise the entire network would simply be a linear model). Shown here is the sigmoid (AKA logistic) activation function which translates and constrains values between 0-1, often used in output layers for a binary classification probability prediction. **Note bias weights are left out for simplification.

A

B

Figure 3. Convolutional Filters and Activations

1. Simplified schematic of a 3x3 convolutional filter (red box) sliding across an image to produce a feature map. At each position, the resulting feature is the sum of the products of the weights of the filter and feature values of the input image. CNN filters act as detectors to extract certain patterns or features from the input – for example this filter acts as a vertical line detector and the resulting feature map is the extraction of the vertical line from the original cross image.
2. Visualization of activations of different convolutional layers provides intuition into the hierarchical nature of learned representations in CNNs. Each small square represents the activation map from an individual CNN filter using a 4-chamber image from an echocardiogram as input. Earlier layers detect edges and textures of the image, while intermediate layers begin to detect more complex features of the image such as ventricular contours or cardiac valves. Later layers prior to the model output become increasingly abstract and difficult to interpret visually as complex concepts (e.g. LV shape) are represented that are more relevant to the specific output of the neural network (e.g. view classification).

Figure 4. Gradient Descent Intuition

Gradient descent is a first order (i.e. using the first derivative), iterative local mathematical optimization algorithm for minimizing the loss function of a ML model. In order to simplify the intuition into gradient descent, one can imagine a hiker on a mountaintop with heavy cloudcover below such that the hiker can only see a few feet in front of him/her (A). If the goal of the hiker is to descend to the base of the mountain as efficiently as possible, he/she will take a step down the steepest visible portion of the mountain. In the same way, the machine calculates the slope of the loss function at the current state of the model (with current parameter values), then adjusts model parameters to take a step down this slope in an iterative fashion to minimize the loss (B).