Artificial Intelligence Algorithm Development for Biomedical Imaging

van Ooijen, Peter; Oever, L.B. van den

Published in:
Artificial Intelligence in Cardiothoracic Imaging

DOI:
10.1007/978-3-030-92087-6_3

Artificial Intelligence Algorithm Development for Biomedical Imaging

3

Peter M. A. van Ooijen and Leonardus B. van den Oever

Introduction

The development of artificial intelligence (AI) in imaging gained momentum with the ability to use deep learning for image classification purposes in the image challenge that was won by Geoff Hinton and his team [1]. From the moment of this feat, medical imaging became one of the prominent areas of research of artificial intelligence and deep learning up to a point where the same Geoff Hinton claimed in a statement in November 2016 at the Machine Learning and Market for Intelligence Conference in Toronto that "we should stop training radiologists now" because they would be replaced by deep learning in a couple of years' time [2]. Since then, a lot of progress has been made, but we are far from replacing medical doctors with AI and the sentiment has changed to the view of AI as an addition to the human expert or augmented intelligence instead of artificial intelligence.

Traditional medical imaging tasks that were supported by conventional image processing software are the segmentation of anatomy or pathology, the detection of pathology, and the classification of pathology. Artificial intelligence started to play its role in these domains but also opened possibilities that were not seen before, such as the generation of synthetic images and the implementation of image data-based predictive algorithms for preventive medicine and treatment planning.

The fast and diverse developments in both the hardware required and the software supporting the development and deployment of AI algorithms create an environment in which the capabilities of AI flourish. However, actual clinical application is still not widespread. Besides ethical and legal reasons, which are enforced because of the weight that medical decisions often carry, this also has technical reasons. The "black-box" nature of deep learning makes it difficult to reproduce and/or explain experimental results and, therefore, renders AI based decision-making incomprehensible for both physicians and patients. This lack of transparency and explainability is also often regarded the main reason that trust in AI among its potential users is still low.

However, it is highly probable that this will change in the coming years when repeated validation of designed AI models will result in more robust and explainable AI implementations, thereby increasing the trust in AI. That hard work is being done to get to that stage as soon as possible can be seen from the sharp increase in the number of papers published on AI, machine learning, and deep learning in cardiology and cardiovascular imaging in the past decade [3]. These papers cover a wide variety of application areas of AI in cardiovascular imaging, ranging from acquisition and reconstruction to prediction and reporting and spanning the full range of data acquisition devices used in cardiovascular imaging from echocardiography [4, 5] to computed tomography (CT) and magnetic resonance imaging (MRI) [6]. With this, it should be noted that all of these currently developed AI systems are so-called narrow AI solutions, or in a more popular term, they are "one-trick ponies." This means that the AI systems developed nowadays are only good in performing one specific task in a very controlled environment and will fail when, for example, unknown new data or data of different quality is presented as input.

Based on the papers published, we can observe reported performance gradually increasing to a level where it is equal to or even better than the human expert. Therefore, it seems that we are on the brink of a more widespread implementation of AI-based tools and many publications have discussed and demonstrated this AI future of cardiovascular imaging [3, 6, 7]. To help get an overview of these rising possibilities and what is technically required to achieve them, this chapter provides insight in the different tasks that can be performed by AI tools based on cardiovascular imaging data and gives an overview of the steps involved in the AI development process.

P. M. A. van Ooijen, MSc, PhD, CPHIMS (✉)
L. B. van den Oever, MSc
Department of Radiation Oncology and Data Science Center in Health, University Medical Center Groningen, Groningen, The Netherlands
e-mail: p.m.a.van.ooijen@umcg.nl; l.b.van.den.oever@umcg.nl

C. N. De Cecco et al. (eds.), *Artificial Intelligence in Cardiothoracic Imaging*, Contemporary Medical Imaging,
https://doi.org/10.1007/978-3-030-92087-6_3

AI in Medical Imaging

AI can have impact on all steps in the medical imaging process from image acquisition to predictive medicine (Fig. 3.1) to provide additional capabilities, automation of tasks, and augmentation of the human operator by providing automated support in a user-guided task.

Image Acquisition

Part of the image acquisition process is the determination of the image quality of the data produced. On the one hand, this is important to the AI algorithms using the data; on the other hand, AI algorithms could also be employed to determine the image quality. This utilization of AI algorithms to assess image quality is especially important in imaging modalities that can have a high variety in image accuracy and quality depending on the equipment used and/or have a high dependency on user experience and skills such as echocardiography. In echocardiography, AI-based image quality assessment could be implemented and used during the actual image acquisition process in real time to guide the user to obtain consistent images of high quality of the correct anatomical positions independent of the experience and skill of the user [4, 8].

Another step in the process of echocardiography acquisition that could benefit from AI is view classification. In clinical practice, the classification of the plane of recorded echocardiography videos is done manually and susceptible to misreporting. Therefore, automatic classification could reduce the workload of the analysts and reduce mistakes. This automation can be achieved using convolutional neural networks (CNN) with cluster analysis to sort input images into five predetermined standard views (e.g., standard cardiac MR views such as short axis, long axis, 2 chamber, 3 chamber, 4 chamber), with a reported accuracy for such classification models of 91–98% [4, 9, 10].

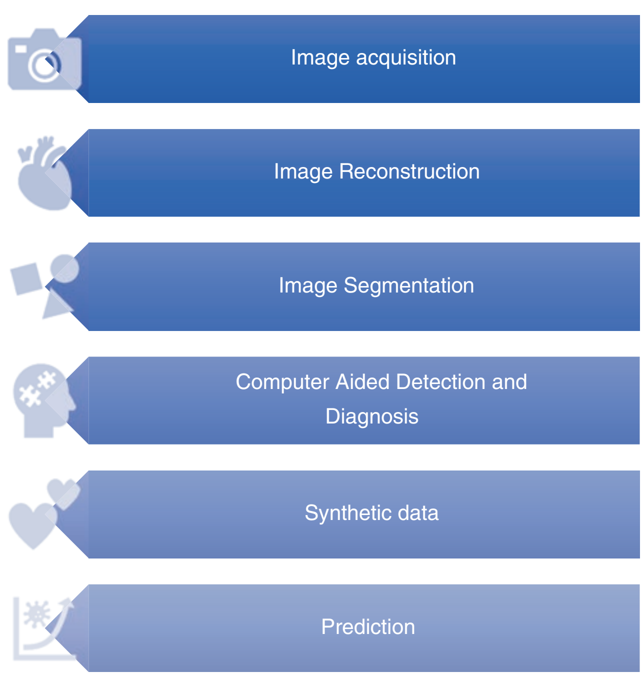

Fig. 3.1 Different possible applications of AI in cardiovascular imaging

In combination, these kinds of models could be applied to guide the user in obtaining the right images for a certain clinical question, thus allowing less experienced operators to acquire high-quality data of the different views.

Another application of AI in image acquisition is the guidance of the technologist in the acquisition of CT or MR examinations to automatically target the right field of view with optimal settings for the individual patient.

Image Reconstruction

Noise reduction is one of the applications of AI during the image reconstruction phase. To achieve noise reduction, a deep learning network is trained – for example – to transform (ultra)low-dose CT data into data that closely resembles the image quality of a full-dose diagnostic CT. This would allow to decrease dose even further and still obtaining diagnostic images. Both convolutional neural networks (CNNs) and generative adversarial networks (GANs) are reported to be able to perform such noise reduction without deterioration of the structure of the image and retaining the diagnostic value of the images [11, 12].

Another benefit of deep learning in the phase of image reconstruction is the ability to decrease the time required to perform cardiac MRI (CMR) examinations by undersampling data collection and using deep learning to estimate the sparse domain based on existing data [13, 14]. Like many AI applications, this method requires large amounts of data to properly train, something that has often been an issue in CMR-related research. Current work, therefore, is restricted to only very specific MR sequences.

Image Segmentation

The most popular application of AI in cardiovascular imaging is undoubtedly the segmentation of cardiac structures in the acquired imaging data, and it is applied to echocardiography, MRI, and cardiac CT in over 100 publications to segment a variety of anatomical and pathological structures [14]. According to the current knowledge base of published work, there is not one specific AI methodology most favorable for the segmentation task. Selected algorithms depend on factors like the imaging modality, protocol used, and cardiac structure to be segmented [15].

One of the possible implementations of automated image segmentation in cardiovascular disease is the segmentation of the cardiac chambers. Several imaging methodologies such as echocardiography, CT, and MRI can employ AI to perform cardiac chamber segmentation, and based on these segmentations, assessment of dynamic parameters such as left ventricular function and size can be automated. High performance is reported on echocardiography using CNNs with an accuracy of 84% [16]. In comparison, left ventricle (LV) segmentation on CT with CNN reported a Dice similarity coefficient of 0.85 with a mean absolute surface distance of 1.1 mm [17]. Most of the work on segmentation of the cardiac chambers using CNN was, however, done in MRI with a variety of techniques. Dice coefficients of 0.94 (Hausdorff distance, 3.5 mm; average symmetric surface distance (ASSD), 0.7 mm) and 0.95 (average perpendicular distance, 1.81 mm) using CNN [18] and CNN with stacked autoencoders [19], respectively, were achieved. Other publications on CNN [20, 21] and deep belief network combining deep learning and level sets [22] also report good performance resembling manual expert segmentation.

Other publications have shown the applicability of AI on segmentation of cardiac valves [23], coronary calcifications [24, 25], and coronary arteries [26, 27].

Computer-Aided Detection and Diagnosis

Classification into multiple categories is a task that can be performed in a supervised or unsupervised manner depending on the task at hand and the available dataset. In a supervised method, both the source data and the corresponding annotations/labels that are the required outcome or endpoint are available. With unsupervised training, only the source data is used to determine clusters of datapoints with similar properties.

Although most of the developed AI applications employ the supervised method, some demonstrations of successful unsupervised methods are available in the literature. For example, Cikes et al. used an unsupervised clustering approach based on a set of parameters to identify patients with heart failure that benefit from cardiac resynchronization therapy [28], and Shah et al. used unsupervised phenomapping to diagnose heart failure with preserved ejection fraction with an AUC of 0.70–0.76 [29].

Supervised AI was employed to determine myocardial disease from a variety of imaging modalities based on quantification. Examples of such applications using echocardiography imaging are the automatic determination of the left ventricular ejection fraction using CNN with high accuracy of 81–95% [4, 16, 30–32], the detection of mitral regurgitation with accuracy >99% using a support vector machine (SVM) [33], and myocardial infarction diagnosis with accuracy ranging from 87% [34] to 99.5% [35]. Wall motion abnormalities could be classified using CNN with a reported AUC of 0.97 [36] and accuracy of 75% [37] and using conventional machine learning classifiers with an accuracy of 96% [38].

Additionally, in cardiac MR, projects on the determination and calculation of cardiac function parameters have reported high accuracy and low mean absolute difference ranging from 3 to 8.5 mL between manual and automatic measurement of the left ventricular end-diastolic volume, end-systolic volume, stroke volume, ejection fraction, and mass [39, 40].

Where MR and echocardiography have their focus on the functional parameters of the heart, CT has mainly been applied to explore the presence and extent of coronary plaque or calcium and the presence and severity of stenotic lesions. For example, for determination of the coronary calcium score, Wolterink et al. described a method where candidates were extracted by intensity-based thresholding and described by location features derived from estimated coronary artery positions, as well as size, shape, and intensity features. Next, a two-class classifier distinguished between coronary calcifications and negatives or a multiclass classifier labeled CAC per coronary artery. Candidates that could not be labeled with high certainty were identified by entropy-based ambiguity detection and presented to an expert for review and possible relabeling with an intra-class correlation coefficient of 0.95 [41].

Both machine learning techniques such as boosted ensemble algorithms [42] and CNNs [43] were used for the quantitative analysis of coronary stenosis and plaque measurement with reported accuracies of >80%. Another application utilizing machine learning is the determination of fractional flow reserve (FFR) based on CT, resulting in an accuracy of up to 83% [44, 45].

Synthetic Data Generation

AI also allows to use available (imaging) data to generate new data, so-called synthetic data. This ability to create synthetic data got a lot of public attention because of demonstrations from Deepfake and sites like http://thispersondoesnotexist.com but so far has not proven to be very popular to apply in medical imaging research. A technique that is commonly used to construct synthetic data is the generative adversarial network (GAN) as proposed by Goodfellow et al. in 2014 [46]. A GAN makes use of two neural networks that compete against each other, a generative and a discriminative network. The generative network generates an image based on a random input, and the discriminative network tries to distinguish between real data and the generated fake. When the discriminative network is unable to make the distinction, the result is a successful fake (or synthetic) image. The error or loss of the discriminative network is, therefore, the reverse loss of the generative network. The higher the loss of the discriminative network, the lower the loss of the generative network, hence the name adversarial network. The generative network can then be used to generate more synthetic data.

For cardiovascular imaging applications, different varieties of synthetic data can be constructed that could prove useful in clinical practice.

First, synthetic data can be generated based on an available database to increase the number of samples that can be used for training a deep learning network. For example, Diller et al. used a dataset of 303 cardiac MRI scans of tetralogy of Fallot patients to generate 100,000 new synthetic images that were then used as a training set to train a U-net segmentation [47]. They found that on visual inspection, all synthetic samples were classified as anatomically plausible by human observers and that the U-net trained on synthetic data performed comparable to the same U-net architecture trained on real patient data. This shows that the use of synthetic data to extend a training set in order to train a deep learning network is a feasible solution when the amount of available data is limited.

A second application of synthetic data generation is the transition from data acquired on one imaging modality to a new synthetic dataset that looks like it was acquired on another imaging modality. Common transitions under investigation in different domains are to synthesize CT data from an acquired MRI or to synthesize CT data from an acquired cone beam CT. Applications that could benefit from synthetic data generation are MR-only radiotherapy planning or synthetic CT based on cone beam CT (CBCT) obtained during radiation therapy in order to enable adaptive treatment planning [48]. Direct applications of this type of data conversions in cardiovascular imaging are not yet found.

Another possibility of using generative networks is synthetic contrast enhancement, where contrast media enhancement is synthetically added to a non-contrast enhanced scan. Although not many papers exist on this particular topic, it shows promise for some clinical applications and could be valuable for patients that are allergic to the contrast agent or to further reduce the amount or concentration of contrast media administered. Santini et al. showed synthetic contrast enhancement in cardiac CT with deep learning to be feasible and achieved a high Dice score of 0.88 ± 0.03 and a low volume percentage error of 9.1 ± 6.2% for segmentation of the synthetically enhanced left ventricle [49]. However, they only used a small dataset and limited themselves to a 2D representation of the long axis view of the cardiac chambers, thus not providing actual 3D information.

Prediction

Predictive algorithms can be applied to implement early prediction of disease in order to start preventive treatment or to determine future treatment response to allow patient-specific treatment selection.

One concrete example of the prediction possibilities is the prediction of major adverse cardiac events (MACE) based on a combination of clinical, stress-testing, and imaging-derived variables. This kind of prediction relies on the discovery of features in a multi-layered model [6].

Motwani et al. performed prediction of all-cause mortality among patients with suspected coronary artery disease by employing ML techniques (boosted ensemble algorithm) in cardiac CT achieving an AUC of 0.79 [50]. Comparable results with an AUC of 0.771 were achieved by van Rosendael et al. also using a boosted ensemble algorithm (XGBoost) on CT data to predict risk score for all-cause death and non-fatal myocardial infarction during a >3-year follow-up [51]. Based on MRI, Dawes et al. reported an AUC of 0.73 for predictive modeling of survival in patients with pulmonary hypertension using principal component analysis [52]. Also using MRI, Samad et al. performed predictive modeling of deterioration of left ventricular function in patients with a repaired tetralogy of Fallot using a SVM classifier, resulting in an AUC of 0.82 for predicting any deterioration [53].

It is common in ML techniques that use clinical or image-derived variables to first reduce the number of variables used for the targeted prediction. The higher the number of input variables, the more difficult it becomes for an ML technique to train correctly and achieve high accuracy. These dimensionality reduction techniques use one of two basic techniques, feature elimination or feature extraction. The aforementioned XGBoost, a feature elimination method, tests the input variables for their impact on the accuracy of the ML algorithm and discards variables that have little impact on the prediction results, making the ML algorithm more efficient and easier to train. Feature extraction creates new variables by combining the old variables and then dropping all the old variables and the new combined variables that do not add to the prediction accuracy.

AI Development Process

For all of the previously described AI developments, the process is quite similar regardless of whether developing a tool for the technician supporting image acquisition or one for a radiologist or cardiologist to provide outcome prediction (Fig. 3.2). However, differences do exist in certain steps because of the inherent difference in the different methodologies available to the data scientist or AI engineer.

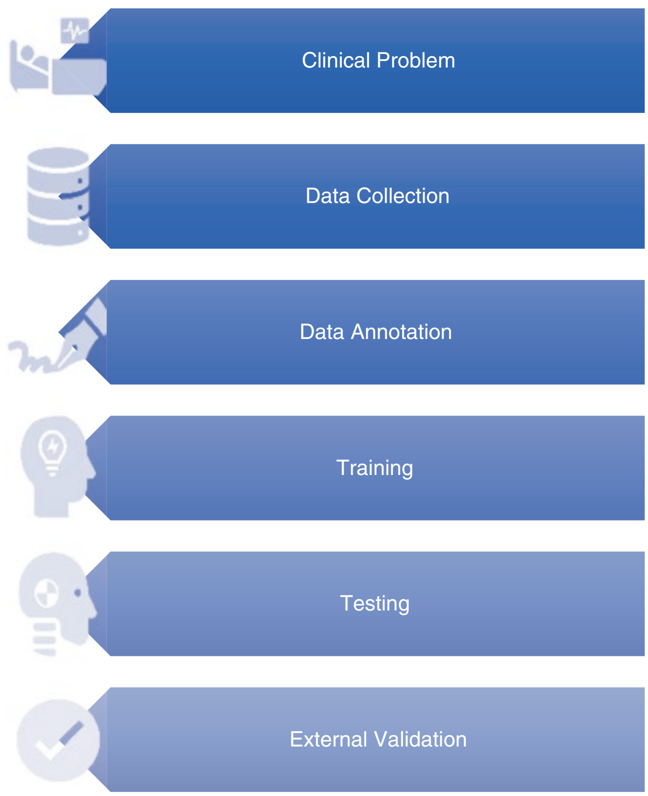

Fig. 3.2 Steps in the AI development process

Machine learning or deep learning methodologies can roughly be divided into three groups, being supervised learning, unsupervised learning, and reinforcement learning.

In *supervised learning*, the system learns by knowing the truth for the training dataset and using the difference between the truth and the prediction (based on the so-called loss function) to iteratively improve the prediction. The known "truth" that is connected to the data can be in the form of annotations or labels to the image. Annotation involves the actual selection of the region of interest on the image itself. This is often used for segmentation tasks. The selection can vary from just identifying the region of interest by placement of a rough bounding box to the full (manual) segmentation of the region of interest by detailed contour drawing. Labeling involves the linking of a textual or numerical label to the image without annotation on the image itself. Such a label could, for example, be a diagnosis (e.g., the existence of a certain disease visible in the image) or a number (e.g., the number of abnormalities shown in the image). Another distinction that can be made in the labeling is a simple binary (yes/no) labeling or a multi-class labeling describing a number of different outcomes. The same holds for annotations where a single region of interest can be available indicating one region with specific properties, or multiple regions can be selected each with their own specific properties. In supervised learning, there are two kinds of learning tasks: classification and regression. Classification models try to predict distinct classes, while regression models predict numerical values.

In *unsupervised learning*, the system tries to discover the hidden structure of data or associations between variables. In that case, training data consists of instances without any corresponding labels or annotations. The associations between variables are defined in the form of clusters. Clusters are informative patterns occurring through clustering, which means the separation of a whole dataset into groups of data, so that instances belonging to the same group are as similar as possible and instances belonging to different groups differ as much as possible.

The term *reinforcement learning* is a general term given to a family of techniques in which the system attempts to learn through direct interaction with the environment so as to maximize some notion of a cumulative reward. It is important to mention here that the system has no prior knowledge about the behavior of the environment and the only way to find out is through trial and error. Reinforcement learning is mainly applied to autonomous systems due to its independence in relation to its environment but can also play a role in medical imaging by providing systems that learn by doing. In such a system, there is no pre-definition of the labels or annotations of the data, but they are provided by the user while the system is learning in the background. After providing sufficient input to the model to obtain a certain level of confidence, the model will start providing the predictions to the user who can then provide input by correcting mistakes to further train and improve the model. One of the major challenges of such systems in clinical practice is that they keep learning and thus change behavior over time, which hampers the legal acceptance and quality assurance of such a model. The learning/training process is dependent on the experience of the user. When a less experienced user is feeding such a system with wrong answers, the system will also learn from this faulty input assuming that it is correct, and eventually the system will adjust itself and start producing answers tailored to the faulty input.

Because of the complex nature, in almost every step of the AI process, the collaboration in a multidisciplinary group of healthcare professionals, data scientists, and deep learning experts is crucial to get to the optimal result.

Clinical Problem

The clinical problem should be the start of every AI project. This problem could be related to the inability of a human observer to retrieve the required information from imaging data, or the time-consuming process of manual processing of images, but also about the inability of the human observer to provide predictions of future events or progression of disease because of the vast amount of data available, which makes it impossible for a human observer to assess it all. All these could lead to the exploration of the support of artificial intelligence to solve the problem. After definition of the clinical problem faced, multiple subsequent steps have to be taken to perform in the AI development process.

Data Collection

The next step to develop an AI algorithm is the collection of a large enough dataset of high quality that represents the target population and holds as little bias as possible. The quality of such a dataset is one of the most important factors of the AI development process and very often the limiting factor. Just as in conventional image processing, the adage "garbage in is garbage out" also holds in the case of artificial intelligence; consistent quality of the data is of utmost importance as is standardization.

When collecting data, it is important to do this based on the clinical question. The data collection should resemble the target population of the eventual tool, contain a large enough portion per subset related to the different endpoints, and have as little bias as possible.

The challenge of bias lies not only in the obvious target population features such as gender, age, and race but also in the data origin. This becomes even more difficult in medical imaging, where the anatomy of patients has a large variation. For training purposes, this means that a dataset needs to include as many samples as possible. In case of limited availability of data, this is often achieved by using either patches from volumes or extracting 2D slices out of volumes, which are often remarkably similar, or by data augmentation during the training phase.

Many of the publications mentioned in this chapter only provide results of an AI system trained, tested, and validated on a single-institution dataset. However, radiological images can vary a lot in their presentation because of the dependency on the imaging equipment used. Because of the bias introduced by training on a database with images of the locally used equipment, acquisition and reconstruction parameters, similar performance is not guaranteed on a database obtained at another institution where these factors differ. For example, Biondetti et al. demonstrated that CNNs can learn to distinguish the scanner manufacturer and that this bias can substantially impact model performance for both classification and segmentation tasks [54].

Data Annotation or Labeling

Besides the data itself, the quality of the annotations or labels connected to the data is also extremely important. The training, test, and validation phase of supervised learning approaches heavily rely on the quality of these data annotations and labeling. Therefore, a number of questions can be asked concerning the annotation and labeling.

A first question is whether previously acquired labels are already available. The previous labels could be obtained from the clinical databases or the reports from a clinical trial or research project. However, this leads to the question whether we can actually retrieve the labels from the clinical or research system used in a usable and readable format. It could therefore also be necessary to specifically (re)label the data because no previous labels exist, or they are not usable for technical or practical reasons. The annotation or labeling is – in many cases – performed by one or multiple human experts and provided as consensus or majority vote of those human experts. This leads to the next question of whether the previously acquired labels are appropriate for the task at hand. This concerns questions about not only the strength of the label but also the consistency of labels that were previously added by multiple people with different methodology or experience. It is important for an ML algorithm to learn a single method of annotating the data. Biases caused by inconsistency of the labels can cause the network to become less accurate. Similarly, for validation and testing, the labels need to be consistent. The ML algorithm will learn to annotate the data in a certain method; if the validation data is then inconsistently labeled, mismatches will negatively influence accuracy. Once we have a labeled or annotated dataset, we also need to consider if there is class balance between the different labels.

In some cases, additional information about the ground truth can be obtained besides the labeling of human observers; if this is possible – for example, pathology results for malignancy prediction – these could be used to strengthen the labels provided.

Training and Testing

The dataset collected will be divided into a train and a test set, often 80% and 20% of the total dataset. This split can be performed in different ways. First, a random split can be performed. However, this will only be sufficient if the dataset is large enough to avoid unwanted bias in either the train or test set. Therefore, a balanced split is also possible where certain factors are balanced over the train and test set, for example, concerning the population (age/gender) or the outcomes (same distribution of the outcomes in both training and test set).

Due to the nature of medical data, most of the times, the classes in a dataset are very imbalanced, usually either toward the healthy prediction or the opposite. This makes the algorithm predictions incline to the majority class, creating poor accuracy on the minority classes that are usually the diseased classes. It should be noted that some quantification methods commonly used will still find high accuracy for these algorithms, since the error will be small if the minority class is also a minority in the test dataset. Oversampling the minority class and undersampling the majority class are techniques commonly employed in the training phase of an ML algorithm to alleviate the class imbalance problem.

The training itself is an iterative process in which a training dataset is provided to the AI model, which will then produce predictions. These predictions are – in case of supervised learning – compared in smaller batches to the true labels, and the difference (or loss) is used to tweak the model to reduce the loss by changing its weights after each batch. After each iteration over the complete training dataset, a separate validation will be done with a part of the dataset that is not used for the training. After *n* iterations (or epochs) over the full dataset, the training will be complete and result in a trained model.

So in this process, the AI model learns by calculating how wrong its predictions are by using the provided correct in- and output.

Many of the deep learning networks are data hungry when training the network and require a lot of data. The number of cases presented to the training phase can be extended by so-called data augmentation. In the case of imaging data, this means that besides the original images, also slightly adjusted images are fed to the deep learning network to expose it to a higher variety of images with different presentations. Possible augmentation adjustments are image mirroring, rotation, zooming, and stretching. This data augmentation is usually performed on the fly during the training process.

External Validation

Sometimes, internal validation is performed by providing a part of the dataset to the model that it never saw before from the own dataset (e.g., by dividing it into 60%/20%/20%). However, a better approach is to obtain a true independent validation by including an external validation set obtained from one or more different institutions. This could be collected, dedicated for this one occasion, but often publicly available datasets are used that are posted on open repositories or were previously released as part of a data science challenge. Using such a set also allows to compare algorithms that report on using the same dataset in their published work.

Multistage Pipelines

Currently, the capabilities of AI software are often not advanced enough for single neural networks to solve the complex problems seen in medical imaging and clinical workflows. Therefore, multistage pipelines of combined neural networks, each for a specific task, are used. One of the examples would be the reduction of variables as discussed in the prediction subchapter. More convoluted problems, such as locating a pneumothorax, will first identify and segment the lungs with a neural network and then train another neural network to classify the segmented image for the presence of a pneumothorax. During development of these kinds of workflows, the pipelines are usually based on clinical workflow. When reading a chest x-ray, intuitively, a reader would identify the lungs and then look for abnormalities. AI does not have intuition, so we have to help it with these steps.

Conclusion

As can be observed from the successful implementations showcased in the growing number of publications, artificial intelligence shows great promise in cardiovascular imaging for different goals [55]. However, most of the studies published are descriptive in nature, and its widespread implementation is still hampered by low availability and reproducibility. Therefore, in the development of AI, careful attention to the different steps in the development process should be taken to avoid suboptimal training of the network. Many challenges still exist in the development of AI tools such as suboptimal training data with bias or incorrect labels, insufficient training and validation, and many more.

However, it can also be observed that AI has made great progress in the past decade and has already made impact in cardiovascular imaging in both acquisition and diagnosis. This shows that AI has the potential to increase the quality of care and reduce the burden on both the healthcare professional and the patient.

References

1. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convoluational neural networks. Adv Neural Inf Process Syst. 2012;25:1–9.
2. [Online]. Available: https://www.youtube.com/watch?v=2HMPRXstSvQ.
3. Verjans J, Leiner T. Artificial intelligence for the general cardiologist. Neth Hear J. 2019;27:389–91.
4. Kusunose K. Steps to use artficial intelligence in echocardiography. J Echocardiogr. 2020;Oct 12 (online ahead of print):1–7.
5. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan C, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R. Fully automated echocardiogram interpretation in clinical practice. Circulation. 2018;138(16):1623–35.
6. Siegersma K, Leiner T, Chew D, Appelman Y, Hofstra L, Verjans J. Artificial intelligence in cardiovascular imaging: state of the art and implications for the imaging cardiologist. Neth Hear J. 2019;27:403–13.

7. Kittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligenve in precision cardiovascular medicine. J Am Coll Cardiol. 2017;69(21):2657–64.
8. Liao Z, Girgis H, Abdi A, Vaseli H, Hetherington J, Rohling R, Gin K, Tsang T, Abolmaesumi P. On modelling label uncertainty in deep neural networks: automatic estimation of intra-observer variability in 2d echocardiography quality assessment. IEEE Trans Med Imaging. 2020;39(6):1868–83.
9. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classifcation of echocardiograms using deep learning. NPJ Digit Med. 2018;1:6.
10. Kusunose K, Haga A, Inoue M, Fukuda D, Yamada H, Sata M. Clinically feasible and accurate view classifcation of echocardiographic images using deep learning. Biomol Ther. 2020;10(5):665.
11. Wisselink HJ, Pelgrim GJ, Rook M, van den Berge M, Slump K, Nagaraj Y, van Ooijen P, Oudkerk M, Vliegenthart R. Potential for dose reduction in CT emphysema densitometry with post-scan noise reduction: a phantom study. Br J Radiol. 2020;93(1105):20181019.
12. Woltering J, Leiner T, Viergever M, Isgum I. Generative adversarial networks for noise reduction in low-dose CT. IEEE Trans Med Imaging. 2017;36(12):2536–45.
13. Bustin A, Fuin N, Botnar RM, Prieto C. From compressed-sensing to artificial intelligence-based cardiac MRI reconstruction. Front Cardiovasc Med. 2020;7:17.
14. Lekadir K, Leiner T, Young AA, Petersen SE. Editorial: current and future role of artificial intelligence in cardiac imaging. Front Cardiovasc Med. 2020;7:137.
15. Dey D, Slomka PJ, Leeson P, Comaniciu D, Shrestha S, Sengupta PP, Marwick TH. Aritificial intelligence in cardiovascular imaging. J Am Coll Cardiol. 2019;73(11):1317–35.
16. Leclerc S, Smistad E, Pedrosa J, Ostvik A, Cervenansky F, Espinosa F, Espeland T, Berg EAR, Jodoin P-M, Grenier T, Lartizien C, D'hooge J, Lovstakken L, Bernard O. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans Med Imaging. 2019;38(9):2198–210.
17. Zreik M, Leiner T, de Vos BD, van Hamersvelt RW, Viergever MA, Isgum I. Automatic segmentation of the left ventricle in cardiac CT angiography using convolutional neural networks. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). Prague; 2016.
18. 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 J. Machine learning derived segmentation of phase velocity encoded cardiovascular magnetic resonance for fully automated aortic flow quantification. J Cardiovasc Magn Reson. 2019;21(1):1–11.
19. Avendi M, Kheradvar A, Jafarkhani H. A combined deeplearning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Med Image Anal. 2016;30:108–19.
20. Tan LK, Liew YM, Lim E, McLaughlin RA. Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. Med Image Anal. 2017;39:78–86.
21. Zheng Q, Delingette H, Duchateau N, Ayache N. 3-D consistent and robust segmentation of cardiac images by deep learning with spatial propagation. IEEE Trans Med Imaging. 2018;37(9):2137–48.
22. Ngo TA, Lu Z, Carneiro G. Combining deep learning and level-set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. Med Image Anal. 2017;35:159–71.
23. Astudillo P, De Beule M, Dambre J, Mortier P. Towards safe and efficient preoperative planning of transcatheter mitral valve interventions. Morphologie. 2019;103(343):139–47.
24. van Assen M, Martin SS, Varga-Szemes A, Rapaka S, Cimen S, Sharma P, Sahbaee P, De Cecco CN, Vliegenthart R, Leonard TJ, Burt JR, Schoepf UJ. Automatic coronary calcium scoring in chest CT using a deep neural network in direct comparison with non-contrast cardiac CT: a validation study. Eur J Radiol. 2020;21(134):109428.
25. Gharaibeh Y, Prabhu D, Kolluru C, Lee J, Zimin V, Bezerra H, Wilson D. Coronary calcification segmentation in intravascular OCT images using deep learning: application to calcification scoring. J Med Imaging (Bellingham). 2019;6(4):045002.
26. Yang S, Kweon J, Roh J-H, Lee J-H, Kang H, Park L-J, Kim DJ, Yang H, Hur J, Kang D-Y, Lee PH, Ahn J-M, Kang S-J, Park D-W, Lee S-W, Kim Y-H, Lee CW. Deep learning segmentation of major vessels in X-ray coronary angiography. Sci Rep. 2019;9(1):16897.
27. Huang W, Huang L, Lin Z, Huang S, Chi Y, Zhou J, Zhang J, Tan R-S, Zhong L. Coronary artery segmentation by deep learning neural networks on computed tomography coronary angiographic images. Annu Int Conf IEEE Eng Med Biol Soc. 2018;2018:608–11.
28. Cikes M, Sanchez-Martinez S, Claggett B, Duchateau N, Piella G, Butakoff C, Pouleur A, Knappe D, Biering-Sørensen T, Kutyifa V, Moss A, Stein K, Solomon S, Bijnens B. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. Eur J Heart Fail. 2018;21:74–85.
29. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, Bonow RO, Huang C-C, Deo RC. Phenomapping for novel classification of heart failure with preserved ejection fraction. Circulation. 2015;131:268–79.
30. Tabassian M, Sunderji I, Erdei T, Sanchez-Martinez S, Degiovanni A, Marino P, Fraser AG, D'hooge J. Diagnosis of heart failure with preserved ejection fraction: machine learning of spatiotemporal variations in left ventricular deformation. J Am Soc Echocardiogr. 2018;31(12):1272–84.
31. Asch FM, Poilvert N, Abraham T, Jankowski M, Cleve J, Adams M, Romano N, Hong H, Mor-Avi V, Martin RP, Lang RM. Automated echocardiographic quantifcation of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert. Circ Cardiovasc Imaging. 2020;12(9):e009303.
32. Kusunose K, Haga A, Yamaguchi N, Abe T, Fukuda D, Yamada H, Harada M, Sata M. Deep learning for assessment of left ventricular ejection fraction from echocardiographic images. J Am Soc Echocardiogr. 2020;33(5):632–5.
33. Moghaddasi H, Nourian S. Automatic assessment of mitral regurgitation severity based on extensive textural features on 3D echocardiography videos. Comput Biol Med. 2016;73:47–55.
34. Tabassian M, Alessandrini M, Herbots L, Mirea O, Pagourelias E, Jasaityte R, Engvall J, De Marchi L, Masetti G, D'hooge J. Machine learning of the spatio-temporal characteristics of echocardiographic deformation curves for infarct classification. Int J Cardiovasc Imaging. 2017;33:1159–67.
35. Sudarshan Vidya K, Ng E, Rajendra Acharya U, Meng Chou S, San Tan R, Ghista DN. Computer-aided diagnosis of myocardial infarction using ultrasound images with DWT, GLCM and HOS methods: a comparative study. Comput Biol Med. 2015;62:86–93.
36. Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M, Sata M. A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. JACC Cardiovasc Imaging. 2020;13(2 Pt 1):374–81.
37. Omar H, Domingos JS, Patra A, Upton R, Leeson P, Noble JA. Quantifcation of cardiac bull's-eye map based on principal strain analysis for myocardial wall motion assessment in stress echocardiography. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). Washington, DC; 2018.
38. Raghavendra U, Fujita H, Gundigar A, Shetty R, Nayak K, Pai U, Samath J, Acharya UR. Automated technique for coronary artery disease characterization and classification using DD-DTDWT in ultrasound images. Biomed Signal Process Control. 2018;40:324–34.
39. 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. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. J Cardiovasc Magn Reson. 2018;20(65):1–12.
40. Suinesiaputra A, Sanghvi MM, Aung N, Paiva JM, Zemrak F, Fung K, Lukaschuk E, Lee AM, Carapella V, Kim YJ, Francis J, Piechnik SK, Neubauer S, Greiser A, Jolly M-P. Fully-automated left ventricular mass and volume MRI analysis in the UK Biobank population cohort: evaluation of initial results. Int J Cardiovasc Imaging. 2018;34(2):281–91.
41. Wolterink JM, Leiner T, Takx RA, Viergever MA, Isgum I. Automatic coronary calcium scoring in non-contrast-enhanced ECG-triggered cardiac CT with ambiguity detection. IEEE Trans Med Imaging. 2015;34(9):1867–78.
42. Dey D, Gaur S, Ovrehus KA, Slomka PJ, Betancur J, Goeller M, Hell MM, Gransar H, Berman DS, Achenbach S, Botker HE, Jensen JM, Lassen JF, Norgaard BL. Integrated prediction of lesion-specific ischaemia from quantitative coronary CT angiography using machine learning: a multicentre study. Eur Radiol. 2018;28(6):2655–64.
43. Zreik M, van Hamersvelt RW, Wolterink JM, Leiner T, Viergever MA, Isgum I. A recurrent CNN for automatic detection and classification of coronary artery plaque and stenosis in coronary CT angiography. IEEE Trans Med Imaging. 2019;38(7):1588–98.
44. 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. J Appl Physiol. 2016;121(1):42–52.
45. Coenen A, Kim Y-H, 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 D. Diagnostic accuracy of a machine-learning approach to coronary computed tomographic angiography-based fractional flow reserve result from the MACHINE Consortium. Circ Cardiovasc Imaging. 2018;11(6):e007217.
46. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adverserial networks. 10 June 2014. [Online]. Available: https://arxiv.org/abs/1406.2661. Accessed 17 Nov 2020.
47. Diller G-P, Vahle J, Radke R, Benesch Vidal ML, Fischer AJ, Bauer UM, Sarikouch S, Berger F, Beerbaum P, Baumgartner H, Orwat S. Utility of deep learning networks for the generation of artificial cardiac magnetic resonance images in congenital heart disease. BMC Med Imaging. 2020;20:113.
48. Chen L, Liang X, Shen C, Jiang S, Wang J. Synthetic CT generation from CBCT images via deep learning. Med Phys. 2019;47(3):1115–25.
49. Santini G, Zumbo LM, Martini N, Valvano G, Leo A, Ripoli A, Avogliero F, Chiappino D, Della Latta D. Synthetic contrast enhancement in cardiac CT with deep learning. 2 July 2019. [Online]. Available: https://arxiv.org/pdf/1807.01779.pdf. Accessed 17 Nov 2020.
50. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah M, Adreini D, Budoff MK, Cademartiri F, Callister TQ, Chang H-J, Chinnaiyan K, Chow BJ, Cury R, Dey D. Machine learning for prediction of all-causemortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. Eur Heart J. 2017;38(7):500–7.
51. van Rosendael AR, Maliakal G, Kolli KK, Beecy A, Al'Aref SJ, Dwivedi A, Singh G, Panday M, Kumar A, Ma X, Achenbach S, Al-Mallah MH, Andreini D, Bax JJ, Berman DS. Maximization of the usage of coronary CTA derived plaque information using a machine learning based algorithm to improve risk stratification; insights from the CONFIRM registry. J Cardiovasc Comput Tomogr. 2018;12(3):204–9.
52. Dawes TJ, de Marvao A, Shi W, Fletcher T, Watson GM, Wharton J, Rhodes CJ, Howard LS, Gibbs JSR, Rueckert D, Cook SA, Wilkins MR, O'Regan DP. Machine learning of three-dimensional right ventricular motion enables outcome prediction in pulmonary hypertension: a cardiac MR imaging study. Radiology. 2017;283(2):381–90.
53. Samad MD, Whener GJ, Arbabshirani MR, Jing L, Powell AJ, Geva T, Haggerty CM, Fornwalt BK. Predicting deterioration of ventricular function in patients with repaired tetralogy of Fallot using machine learning. Eur Heart J Cardiovasc Imaging. 2018;19(7):730–8.
54. Biondetti GP, Gauriau R, Bridge CP, Lu C, Andriole KP. "Name that manufacturer". Relating image acquisition bias with taks complexity when training deep learning models: experiments on head CT. 19 August 2020. [Online]. Available: https://arxiv.org/abs/2008.08525. Accessed 20 Nov 2020.
55. Benjamins J, Hendriks T, Knuuti J, Juarez-Orozco L, van der Harst P. A primer in artificial intelligence in cardiovascular medicine. Neth Heart J. 2019;27:392–402.