REVIEW

A cardiologist's guide to machine learning in cardiovascular disease prognosis prediction

Karl-Patrik Kresoja[1,2] · Matthias Unterhuber[1,2] · Rolf Wachter[3,4,5] · Holger Thiele[1,2] · Philipp Lurz[1,2]

Abstract

A modern-day physician is faced with a vast abundance of clinical and scientific data, by far surpassing the capabilities of the human mind. Until the last decade, advances in data availability have not been accompanied by analytical approaches. The advent of machine learning (ML) algorithms might improve the interpretation of complex data and should help to translate the near endless amount of data into clinical decision-making. ML has become part of our everyday practice and might even further change modern-day medicine. It is important to acknowledge the role of ML in prognosis prediction of cardiovascular disease. The present review aims on preparing the modern physician and researcher for the challenges that ML might bring, explaining basic concepts but also caveats that might arise when using these methods. Further, a brief overview of current established classical and emerging concepts of ML disease prediction in the fields of omics, imaging and basic science is presented.

Keywords Machine learning · Atherosclerosis · Heart failure · Arrhythmia · Genetics · Artificial intelligence

Introduction

A modern-day physician is confronted with a staggering increase of health care data created every day, surpassing the computational capabilities of the human brain by far [63]. While traditionally it has been the art of physicians

Karl-Patrik Kresoja, Matthias Unterhuber, Holger Thiele and Philipp Lurz should be considered as equal first and senior authors.

✉ Holger Thiele
  holger.thiele@medizin.uni-leipzig.de

✉ Philipp Lurz
  Philipp.Lurz@gmx.de

1   Department of Internal Medicine/Cardiology, Heart Center Leipzig at University of Leipzig, Struempellstr. 39, 04289 Leipzig, Germany

2   Leipzig Heart Institute, Leipzig Heart Science at Heart Center Leipzig, Leipzig, Germany

3   Department of Cardiology, University Hospital Leipzig, Leipzig, Germany

4   Clinic for Cardiology and Pneumology, University Medicine Göttingen, Göttingen, Germany

5   German Cardiovascular Research Center (DZHK), Partner Site Göttingen, Göttingen, Germany

to incorporate available data into clinical decision-making, this task seems to be unsolvable in the reality of the digital century. There is a broad consensus that artificial intelligence (AI) or machine learning (ML), often mistakenly used interchangeably [66], might pose a solution to this pressing issue. At the same time ML is seen as something unexplainable and unreachable by many health care workers. Soon, almost every physician will be using ML technology either consciously or weaved in the backend of medical software [76]. ML will likely aid us in diagnostic and management decisions, especially those that incorporate large digitized data [63]. In line with this, it is important for physicians and scientists alike to have a basic understanding of ML, to allow for a safe and conscious use in daily practice. The first part of this review will focus on providing a definition of ML and its capabilities, as well as limitations, while the second part will bring closer modern-day applications in cardiology including basic and translational science.

Basic concepts of machine learning

The term ML was coined in 1959 by Arthur Samuel and refers to the idea that computer algorithms will learn and adjust and thereby improve automatically within their given

boundaries through the use of data [67]. ML is considered to be a subset of AI and is therefore not an interchangeable term, and using it as synonym should be avoided. While ML and AI can clearly be distinguished from each other, it is quite hard to do so with conventional statistical methods [13]. Methods like hierarchical- or k-means clustering [31, 70] or linear regression [48] are increasingly termed as ML approaches. In fact, it is not easy to draw a clear line between conventional statistical approaches and ML [13]. In the absence of a uniform discrimination, both approaches can possibly be split according to their purpose: While conventional statistical models try to inform on relationships between variables in a mostly linear manner, ML is mainly focused on providing optimal predictions—often sacrificing interpretability [13]. There is no clear line to draw to distinguish conventional statistical models and ML. ML models for example can provide various degrees of interpretability ranging from the highly interpretable regularized regression method [56] to impenetrable neural networks, but in general, ML models will sacrifice interpretability for predictive power. In practice, this loss of interpretability is hard to accept, as the human mind is mainly used to understand linear associations between variables [30]. The interest in clinical ML and AI applications is increasing, yet it is important to keep in mind that there currently are no randomized clinical trials, which have shown that ML guided decision-making is superior to conventional approaches. However, ML represents a new chapter in science which has to be addressed, and by adding knowledge of ML methods to the toolbox of scientists and clinicians the black box can be opened and understood and future trials will determine the value of ML in clinical practice [27].

Frequently used machine learning approaches

ML models can roughly be split into unsupervised and supervised approaches. While unsupervised ML focuses on discovering connections among variables, supervised approaches aim to achieve the optimal prediction of labeled samples (e.g. patients experiencing or not experiencing a defined outcome). Using unsupervised models, a classifier learns to infer relationships within the given data (e.g. to identify some clusters of patients who may carry the same genetic features of risk factors). Therefore, unsupervised ML can be of use to find new associations and suggest new hypotheses for new study designs, but its capacities in terms of predicting future events are limited. As this review focuses on prognosis prediction, we will focus the manuscript on supervised approaches and methodical concepts. Examples of unsupervised ML in cardiovascular medicine have extensively been reported before [25, 27, 32, 40, 44, 51, 63, 69–72].

Linear and logistic regression likely represent the most well-known statistical methods that can be referred to as ML approaches. They are hindered by a large number of assumptions, which are often violated in medical literature [22]. These include linearity, normality, homoscedasticity, as well as independence of data [82]. When these statistical assumptions are met, those models tend to perform well on training data but might struggle to make accurate predictions in never seen data, especially when linear assumptions are violated (Fig. 1). They tend to have low bias but high variance. For easier interpretation, a high
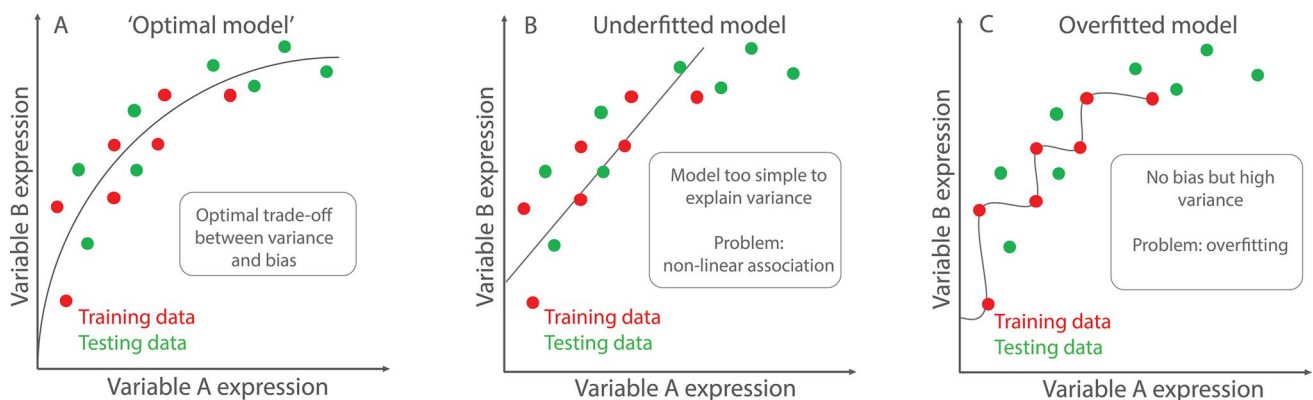

Fig. 1 Examples of difficulties which might occur when choosing unfitting models. The figures represent examples of problems of under- (**B**) and overfitting (**C**) models. **A** Shows the example of an 'optimal' model, where an ideal trade-off between variance and bias is achieved. **B** Shows the example of an underfitted model. Albeit the data follow a non-linear relationship, a linear fit for the data was chosen. This often happens as linear regression might be preferred as the simplest solution to statistical problems, without acknowledging the true spread and association of data. **C** Shows the problem of overfitting, where the model is performing excellent on the training data, but struggles to predict unseen (testing) data. The model 'learns' the data rather than generalizable rules. Using machine learning solutions, one is usually rather prone to over- than underfitting, highlighting the importance of using external testing data to verify the generalizability of the model

bias is useful to accommodate human associative thinking and to discover correlations between independent and dependent variables, e.g., a biased model will tend to ignore outliers, creating a simple rule "valid for all". This penalizes events with fine-grained influences which are unlikely to happen. A model with high variance starts from the assumption that relationships are complex and allows a model to assume many divergent predicting points. This will include more or less outliers for a certain degree, in turn it will not create straight lines, losing generalizability and interpretability. Further, this approach makes the model vulnerable to overfitting.

A possible way to improve the generalizability of a model is by performing so called regularized regression, which introduces a defined error termed λ. This penalization causes the model to perform a little bit worse on the training data but improves its prediction capabilities on unseen data [27]. The goal is to reach a favorable trade-off between the models' decrease in accuracy and improvement in generalizability. Lasso regression can further be used to remove variables which add little to the overall model, which decreases the model's complexity [15].

Supervised ML algorithms see a rapid increase in medical sciences as linear algorithms are often not capable to fully replicate biological reality [78]. For example, specific drug–drug interactions only might occur depending on the state of another variable [45]. Traditional statistical methods are limited in modeling such interactions, especially when considering a larger number of variables. One example are decision tree approaches which often can outperform traditional linear regression analysis on derivation, but due care must be given to not overfit models which will result in poor performance on validation data [18, 78]. Decision tree algorithms such as random forests [80], extreme gradient boosting or its predecessors adaptive boosting [83] are based on simple decision trees. Given a relevant number of variables, a normal decision tree can make a perfect prediction on a given dataset but will rather 'memorize' the shown dataset than learn generally applicable rules-the model 'overfits' (Fig. 1). The bias of such a model is zero, while the variance will be very high. Therefore, the aim in every ML decision tree algorithm is to reduce overfitting to avoid learning noise instead of generalizable patterns. Random forests were one of the first solutions described to address this problem. Instead of a very large and powerful 'zero bias' tree, random forests produce many smaller decision trees with data randomly sampled out and creation of random duplicates of the original dataset ('bootstrapping'). Every produced tree could predict a different outcome for the same patient. Therefore, random forests do not provide a single answer but a 'vote', which comes down to the number of tree classifiers voting for either result this approach is often referred to as 'bagging' [11]. A further development

of decision tree approaches are decision trees leveraging boosting methods. Again, ensembles are constructed from decision tree models, but unlike random forests, individual trees are not built on random subsets of data. Boosting approaches sequentially put more weight on instances with wrong predictions, so they learn from past mistakes. Trees are added one at a time to the ensemble and fitted to correct the prediction errors made by prior models. The latest development of boosting algorithms is the extreme gradient boosting (XGboost) approach recently developed by the University of Washington, which does not only outperform other decision tree approaches in terms of accuracy but has also an eighteen-times faster computing speed than random forests [16].
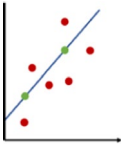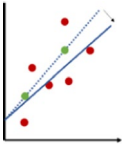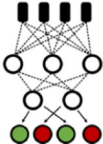
The last supervised ML approach we want to mention is deep learning (DL). DL aims to solve complex problems by mimicking the organization and functionality of the human brain with neural networks [27]. Nodes, which are labeled 'neurons' are arranged in a network layout. The first level of neurons feed into a finite number of other nodes called 'hidden layers' and can be considered as many layers of regressions. When a certain threshold of input is surpassed, a neuron in the hidden layer is 'activated' and by itself passes values further to neurons in the next layer. Like in linear and binomial regressions, these thresholds can be triggered by different values and follow straight, curved or edged lines. This goes on until the final layer is reached which is called output layer. DL excels at analyzing imaging data and are widely used for applications such as facial recognition or image enhancement. The training of DL networks require immense computational capacities, making their processing speed either very slow or very hardware demanding [27]. Reports of use of basic neural networks in cardiology data go as far back as to 1995 [6, 61]. Profound introductions into DL can be found elsewhere [7, 17, 24, 37, 65].

Table 1 shows a summary of advantages and drawbacks of the presented approaches to supervised ML.

Limitations of machine learning

ML is not the panacea for every unsolved and yet to come problem in medical sciences. As any other statistical model, ML models are limited by the quality and magnitude of signal in the dataset from which it is trained. The observation of an event is assumed to be the result of many causal factors. Theoretically, by knowing every predisposing factor and causal variable, an event could be perfectly predictable. In nature, however, adding variables to the equation can lead to noise due to measurement errors and methods, further many outcomes are associated to ever-changing non-predictable environmental factors (e.g., communicable diseases leading to infections triggering cardiovascular events). Even with

Table 1 Examples of machine learning approaches

	Conventional regression	Regularization	Decision trees	Deep learning| neural networks
Examples	Linear regression Logistic regression	Ridge regression LASSO Elastic net	Simple decision trees Bagging Random forest Gradient boosting XGBoost	Convolutional neural networks recurrent neural networks long short-term memory networks
Model accuracy on linear data	↑	↑	↑	↑
Model accuracy on non-linear data	↓↓	↓↓	↑↑	↑↑
Model performance on imaging data	↓	↓	↓	↑↑↑
Interpretability	↑↑↑	↑↑	↓↓[a]	↓↓↓
User know-how required	↓	↔	↑↑	↑↑↑
Processing speed	↑	↑	↑	↓↓
Hardware requirements	↓↓	↓↓	↓	↑↑↑

[a]Interpretability is good for conventional decision tree models but becomes less when using bagging approaches with multiple decision trees incorporating votes on decisions

the 'perfect model', the prediction can only be as good as the true connection between independent and dependent variables but in reality the unpredictability of external cues makes prediction of 'all events and outcomes' impossible [27]. So, while models are not able to provide an absolute truth, they are very good at providing probabilities.

Another practical problem is the quality of data used to train models. While omics analysis provides highly standardized and reproducible data, data from clinical practice might often be prone to higher variance alleviating relationships between outcome and input variables (e.g., NYHA class assessment). Imaging data also seem very standardizable at first thought, but there is a broad variety of different standards with respect to hardware and vendor software, making transferal of imaging data with sufficient quality often challenging. Lastly, there has been a broad discussion on a selection bias introduced into early established ML approaches used for facial image reconstruction due to a lack of diversity in the validation cohorts [49]. This emphasizes that more effort should be put on validation of established ML models in independent cohorts, rather than development of new approaches that only work well in the region or specific subset where it was developed.

Also of importance is the outcome the model seeks to predict. Mortality for example is an excellent outcome to predict as it is an absolute and discrete outcome, with an undoubtable definition. Other factors that are influenced by physiological variance like blood pressure measurements are more difficult to predict. Also, outcomes which require subjective definition might be challenging. While every

endpoint adjudication committee will have no difficulties agreeing on whether a patient is dead, defining 'simple' categories such as cardiovascular death or heart failure hospitalization might become a point of disagreement. Accordingly, a study showed that a ML algorithm struggled to predict 30-day readmission for heart failure (C-index between 0.59 and 0.62) despite being fed with over 200 different variables accounting for demographics, socioeconomic status, medical history, characterization of heart failure (HF), admission and discharge medications, vital signs, weights, selected laboratories treatment, and discharge interventions [23]. Other studies have added to this demonstrating that ML models are better in predicting death than HF hospitalization given the same input data [53]. It is further important to keep in mind that there are outcomes that can be predicted well, but have only limited therapeutic use. In the clinical context, prediction should be focused on cases where actionable consequences for individuals arise.

Lastly, there is a disparity between the kind of answer provided by ML algorithms and the kind of answer required by physicians. Clinicians are expected to make a yes or no decision (e.g., initiate treatment or withhold), while ML algorithms provide probabilities (e.g., 64.73% chance of a patient responding favorably to initiating a specific treatment). A practical example can be found in everyday lab-charts. For example, high-sensitivity cardiac troponin T is usually presented with a cut-off of 14 pg/ml, which represents the upper limit of normal for a healthy reference population [33]. While for the physician it is very clear that a value of 15 pg/ml has a strikingly different

risk than a value of 1400 pg/ml, guidelines have to provide decision-making support and break probabilities down into a 'yes' or 'no'. This leads to a loss of valuable information and is called improper dichotomization [14, 28]. Dichotomization might become even less appropriate for lab values like hemoglobin, which might show U-shaped association with cardiovascular risk [57]. Therefore, physicians should keep in mind that most dichotomizations carry a significant loss of information and an informed physician will treat patients according to their personalized risk, as many already intuitively do [21, 27].

The present use of machine learning in cardiovascular medicine

Papers on ML have skyrocketed in the last years with over 80,000 Pubmed entries referring to either 'machine learning' or 'artificial intelligence'. Using a miner algorithm kindly provided by Quer et al. [63] we displayed current uses of ML with respect to the source data used, the specific disease field investigated as well as the modes of ML used (Fig. 2). As shown, atherosclerosis is the main domain of ML use. Most of the input data used come from imaging like computed tomography (CT)- or magnetic resonance imaging (MRI) scans. DL makes up the largest fraction of ML models, which again is not surprising as they excel at interpreting these kind of imaging data [84]. Figure 3 shows a heatmap of the number of papers on ML in accordance to the subspeciality as well as input data used. Notably, basic science shows only a minor use of ML. In basic science understanding associations between variables might be more important than establishing strong but uninterpretable prediction models, yet there might be some promising new ML opportunities in this field discussed below.
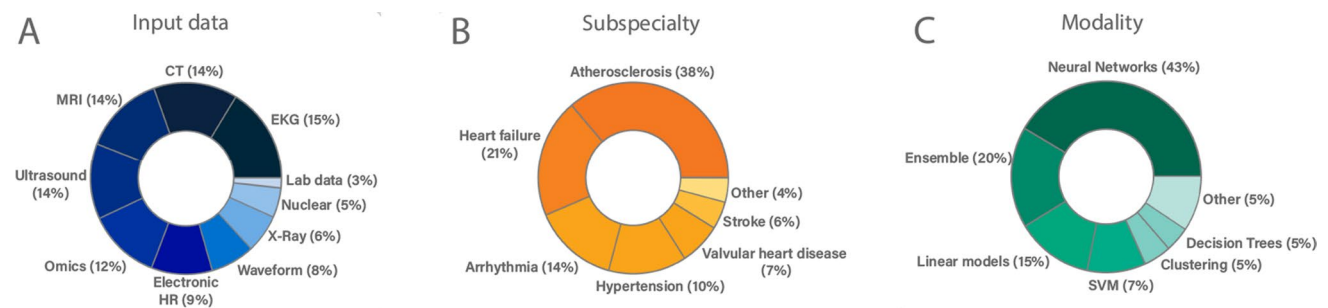

Fig. 3 Heatmap of machine learning use in accordance to subspeciality by input data type. Heat map indicating the number of manuscripts with respect to disease and machine learning modality across the cardiovascular field. Data were mined by an algorithm kindly provided by Quer et al. [63]

Prognostic value of machine learning in omics

The development of high-throughput platforms has left clinicians and scientists with a dilemma. It is now possible to analyze thousands of proteins, metabolites, genes, etc. with a minimum of material and effort. While statistical solutions have significantly developed over the last decade, they still have been outpaced by the rapid development seen in laboratory analytical approaches. This held true until the recent advent of ML approaches. ML enabled predictive algorithms now allow to better model the intricate working of physiology and pathophysiology alike [64].

Despite a vast number of established clinical risk scores for cardiovascular patients predicting outcomes among patients on an individual level remains challenging [38, 62].

Fig. 2 Current use of machine learning approaches in accordance to publications on Pubmed. Distribution of publications on machine learning according to **A** data types, **B** diseases and **C** ML methods used. Other includes basic scien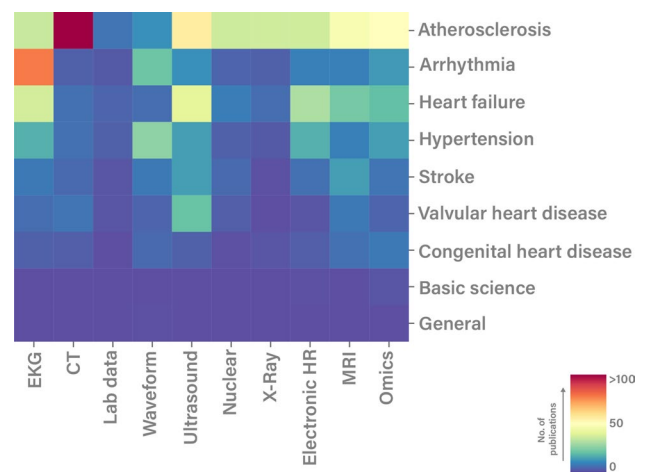ce and congenital heart disease for the subspecialty section B and nearest neighbor method, as well as Gaussian and Bayesian analysis approaches for the modality section C. Data were mined by an algorithm kindly provided by Quer et al. [63]

The incorporation of biomarkers from large omics panels has shown promising results in the prediction of long-term mortality (C-index conventional cox regression 0.65 versus 0.93 for a XGBoost model, net reclassification improvement 78%) [78]. ML does not only improve long-term predictive capabilities but has also shown superiority in short-term prediction of mortality among patients with cardiogenic shock, by using biomarkers [15]. Performing advanced and importantly individualized risk prediction for patients might lead to intensified treatment in high-risk patients and therefore tailor optimal therapy for patients in dire need for optimized care [9]. But not only does ML enable the predictive capabilities of large omics data, it might also improve our understanding of these. By allowing for modeling of complex interactions, proteins are not forced into linear relationships and interpretative approaches like Shap values might be used to acknowledge the non-linearity of biological processes, or to identify possible novel treatment targets [79].

Amongst omics approaches, genetics represent one of the most growing fields in cardiovascular medicine and shows huge potential. Due to the goal of understanding pathways in disease-causing genetic disorders, ML has gained importance and is commonly used in genome-wide association studies (GWAS) [1]. ML methods have been successfully applied to predict the incidence of hypertension by using polygenic risk factors [39, 41, 58], to predict advanced coronary calcium [60], inheritable cardiac disease[12] and to predict type II diabetes in a multi-ethnic cohort [47]. Of interest, the number of layers within neural network architectures used in genomics has generally been far less than those used for image recognition [77], and typically consist of only a few layers [86] with many hundreds to thousands of parameters [35].

Another approach where ML might pave the way for cardiovascular medicine is the so called "liquid biopsy". Liquid biopsy is a minimally invasive technology for detecting molecular biomarkers of a tumor without an invasive biopsy and has been established in oncology [8]. Liquid biopsy came up as a non-invasive way to characterize circulating biomarkers of tumor origin. Like in oncology we one day maybe will be able to fully characterize cardiovascular diseases such as heart failure, acute myocarditis, or coronary artery disease, without the need of further invasive testing and the combination of ML methods and multiomics approaches might be the way to achieve this ambitious goal [42].

Prognostic value of machine learning in imaging

The capabilities in classifying objects in the entertainment and leisure sector started to improve with Kaggle competitions [10]. Up until today, image recognition abilities of neural networks are the most widespread application of DL in clinical practice. ML approaches might especially excel when analyzing standardizable two-dimensional images like electrocardiogram (ECG) data. For example, using a previously trained convolutional neural network in a prospective designed, non-randomized trial, ML was able to identify the occurrence of atrial fibrillation among patients at risk for a stroke with an odds ratio of 5.0 (95% confidence interval 2.3–5.4) [59].

For coronary artery disease (CAD), there is a large number of established ML algorithms for interpretation of CT scans and recently even approaches that aid decision-making and interpretation for invasive coronary angiography were proposed [54]. ML does not only aid in identifying patients with CAD but is also useful in prognosis prediction among these patients [2]. Using a mix of readily available clinical data and data from coronary artery CT scans Motwani et al. were able to show that ML outperforms conventional statistical models with regards to 5-year mortality prediction [55]. A combination of ML methods was presented within a prospective study of 1.912 patients with coronary artery CT scans, where extracardiac adipose tissue was quantified using a fully automated DL approach. Those data together with clinical variables, plasma lipid panel measurements, risk factors, coronary artery and aortic calcium scoring were analyzed by an extreme gradient boosting model and showed excellent predictive capability with respect to the occurrence of myocardial infarction and cardiac death [20]. Investigating approaches where different ML methods are combined on multiple levels could pave the way for an individualized patient risk assessment.

ML algorithms are also increasingly used to automatically interpret echocardiographic images and calculate factors like left ventricular ejection fraction to ease diagnosis of HF in clinical practice and have shown to outperform humans [5, 85]. (NCT05140642) These approaches might be leveraged to increase the interpretation as well as processing speed of lab animal echocardiographic assessment, increasing analysis speed while decreasing the need for additional workforce, after a stable workflow has been established.

Prognostic value of machine learning in basic research

ML cannot only be used to predict clinical outcome but can also be leveraged to predict defined responses, or identify patterns previously missed by conventional approaches.

Most modern high-throughput genomics like Hi-C essentially represent a multilabel image classification problem and therefore DL approaches are ideal to address this problem. Recently, a DL framework called EagleC was presented to detect structural variations in human genome data. This

algorithm was able to capture a set of fusion genes that are missed by whole-genome sequencing and was applied successfully to bulk and single-cell genomics in studying structural variation heterogeneity of primary tumors from Hi-C maps [81].

ML might enhance our understanding of biological processes when accurately dissecting them in logical units. Clerx et al. recently presented elegant ion-channel research, where they investigated the effects of mutations in the SCN5A gene on clinical phenotypes. New unclassified variants of SCN5A are regularly found but predicting their pathogenicity has proven exceedingly difficult. Clerx et al. established a two-step ML model which first aims to establish effects from the gene to functional properties (from gene to function) and then adds a second step where predicted functional properties are used to describe a clinical phenotype (from function to phenotype). They were able to show an improvement in the prediction of functional sodium channel properties and outperformed traditional approaches with the prediction of clinical phenotypes, but predictive performance remained limited [19]. The strength was to split the process into a multi-step approach, which is often the case in biological reality, but the approach was limited due to the imbalance of the available data favoring pathogenic mutations. This practical example highlights the importance of quality of both the outcome as well as the input data used for ML models.

DL has also been described to excel in aiding the analysis of cellular imaging. Beneficial use with image classification, segmentation, object tracking and augmenting microscopic images has been demonstrated. DL for these applications is still in its early phase but has shown promising results and are positioned to render difficult analyses routine and to enable researchers to carry out new, previously impossible experiments [52].

Recently, AlphaFold 2, a ML software developed by Alphabets DeepMind company, has been acknowledged as a milestone in the problem of protein folding [29]. As a protein's physiological function is determined by its three-dimensional structure, the knowledge of latter is essential to understand the biological processes involved [68]. Protein folding is hard to forecast and while models based on simulating quantum mechanics showed promising results, calculation of protein folding in large proteins is challenging due to the exponential rise in required computational capacity. Fundamentally, protein folding is an imaging challenge. It was, therefore, speculated that ML, especially DL, might perform very well. In fact, AlphaFold 2 using novel deep learning approaches, has altered the field of protein folding prediction, vastly outperforming all other 145 presented approaches at the CASP14 competition (Critical Assessment of Structure Prediction, a bi-annual competition for the prediction of protein folding) [73]. AlphaFold 2 as an open-source end-to-end user software is expected to have a relevant impact on the field of protein folding prediction making it more time efficient and accessible for public. However, it is important to keep in mind that, while encouraging, AlphaFold 2 has the limitation that it was trained on the protein data bank (PDB), which incorporates many protein structures which were observed only during experimental conditions, which might not reflect biological reality. Therefore, while AlphaFold 2 is a masterpiece of ML programing, as any other model, it is not protected from the bias of the data used to train the model. Ultimately, ML aims at prediction or reflecting reality, with the latter is always difficult to define and derive [46].

Lastly, ML can also be used to inform on pathways involved in outcomes of interest following steps undertaken by specific peptides, all the way up to their origin in the genome [79]. Jaganathan et al. successfully trained a DL algorithm to predict splicing from pre-mRNA out from a genomic sequence to detect noncoding mutations in rare genetic diseases, possibly paving the way for a reverse engineering of the human genome-protein processing [26]. Further, ML was used for designing molecules to modify specific targets, but also to influence properties like solubility and bioactivity, raising the level of this former trial and error approach to a precision process [75]. In the near future, ML approaches might, therefore, allow treatment response prediction, assessment of in silico protein interactions, identification of novel drug targets, monitoring and predicting response to latter; all of these leading towards a patient-tailored precision medicine approach. There are already interesting commercially available systems like IBM Watson Health's cancer AI algorithm, an algorithm used for recommending treatments for patients with cancer trained on simulated cases. However, the algorithm has been shown to give erroneous outputs in some real-life cases underlying the importance of proper input data and prospective validation [76].

How to interpret ML applications as clinician

Cardiovascular medicine is well positioned to leverage ML methods to facilitate precision medicine approaches by integrating the vast abundance of 'omics' and clinical data available (Fig. 4) [4, 76]. Scientists and clinicians alike often reject ML approaches as they argue that they are a 'black box' and the predictions made by ML are not understandable [50]. Yet, in practice we rely on many established scores and predictors which we do not truly understand. In fact, most scientists and clinicians will acknowledge that they do not know the derivation and validation of a large fraction of scores they use. While they might be easy to calculate without the knowledge of how they were derived and validated

the trust we lay upon them is in fact no lesser a 'black box' than the trust we may lay upon ML. It is important to keep in mind that interpretability might mean different things for clinicians, scientists but also patients [36]. Interpretability by itself might also have different levels. Van der Schaar proposed 3 types to provide interpretability for ML models. The first one is to show explanatory patient features. This means highlighting what features were considered in the ML model and how they were weighted. Gain plots and SHapley Additive exPlanations (SHAP) values provide practical example of this application [43]. The second type suggests grouping patients according to similarities with regards to the predicted outcome, this might allow to identify specific high- or low-risk phenotypes. The third type is vastly harder to establish as it suggests identifying rules and laws, that cause the model to significantly alter a decision and therefore reduce the model to the most important decisions. The first and second types are already practical reality and used in many papers, but while they seem to thoroughly inform on the model they might also introduce dangers as they do not allow on the causal relationship of outcome and the variables of interest. The third type which allows the establishment

of rules might likely be the most sought-after approach in clinical practice, but up until today no valid method exists to establish these laws.

To be able to interpret new ML solutions, it is essential for physicians to apply structured criteria. In line with this the U.S. Food and Drug Administration (FDA), as well as the Standards for Reporting of Diagnostic Accuracy Study (STARD) have established standard of 'Good machine learning practice' and STARD-AI guidelines, respectively, to allow for a common scientific standard [74]. In line with this we provide an overview of questions inspired by Meskó et al. [50] that physicians can ask themselves when assessing new ML approaches in Table 2.

Outlook

ML poses a vast amount of opportunities, but this does not come without the cost of certain pitfalls which have been discussed within the first part of this review. In cardiology, ML has already reached center stage and many fields have already seen interesting proof-of-concept studies especially

Fig. 4 The use of machine learning models to integrate complex multi omics data as well as clinical data and their potential to support clinicians and scientists. Machine learning approaches have the potential to integrate large numbers of variables from large populations to allow for individualized risk prediction. This can be translated into clinical practice by reporting on important features, establishing clinical phenotypes with comparable outcomes and ultimately identifying novel pathomechanisms
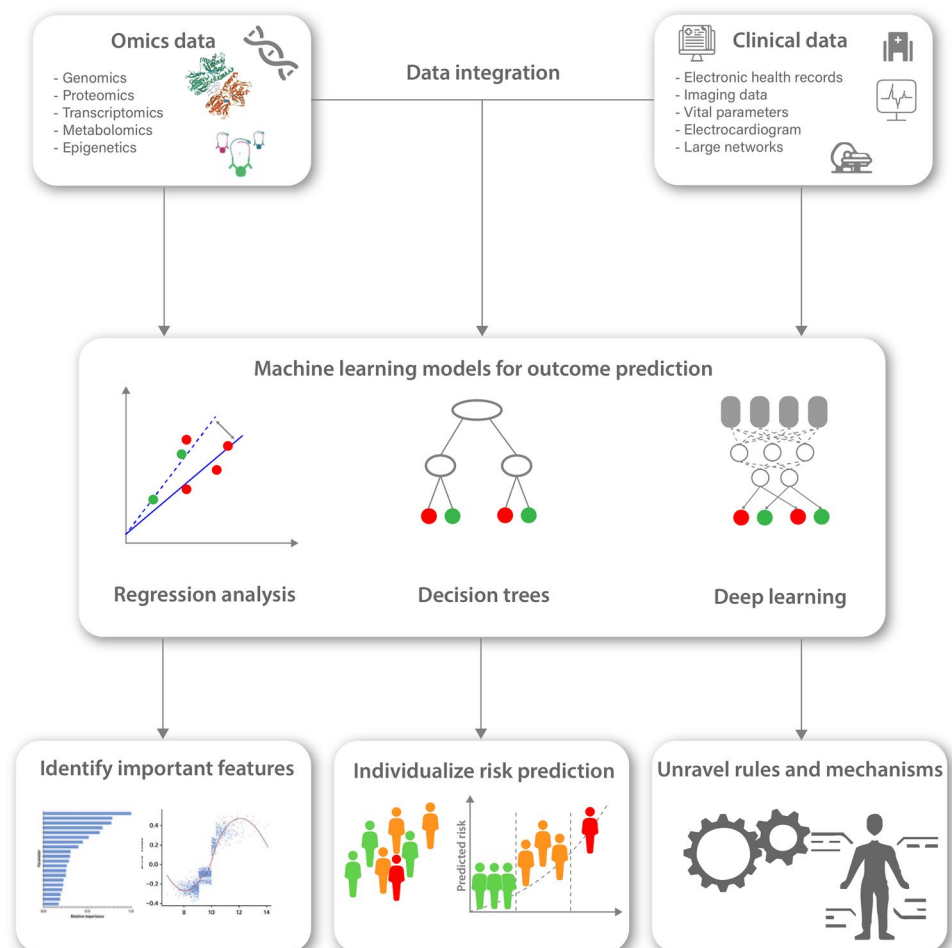
Table 2 General questions to ask when assessing new machine learning models in clinical practice

Issue	Important questions to ask
Relevance of the machine learning model	Does the ML model provide incremental value over established non-ML models?
	Does the predicted outcome have actionable clinical consequences or other relevance? (either for patients and/or relatives)
Sample size and validation	Does the ML model study provide a validation cohort?
	Are the sizes of the derivation/training and validation datasets justified?
	Are there sufficient overall cases to justify a ML approach at all? (e.g., Orphan diseases)
Selection bias	Is the investigated cohort representative for my patient population?
	Do the incidences/events observed reflect real-world numbers?
	Will the model be applied to comparable situations as it was developed in?
Labeling outcomes	How were outcome labels defined?
	Does the labeling used reflected the current gold-standard in the specific field?
	Are the labels standardizable and transferable to other cohorts? (e.g., omics data)
	Are there differences between imaging software/hardware/vendors that hinder easy integration of data sets?
Guideline adherence	Does the ML algorithm development and reporting follow the U.S. Food and Drug Administration (FDA), or the Standards for Reporting of Diagnostic Accuracy Study (STARD) guidelines?
	Was there a prespecified analysis plan published?
Transparency	Is there supporting information on how the ML model derives its decisions?
	Is the ML model publicly available?
	Are there monetary/patent interests of the authors or other groups on the presented ML algorithm?

as medical data is increasing in amount and complexity daily. However, it is important to keep in mind that ML approaches are not always the optimal solution and especially in basic research where causal associations between variables might be more important than optimized predictions, conventional statistical approaches might at this stage provide better use than modern ML techniques. But acknowledging the potential of ML methods which allows non-linear assessment of associations, that further accounts for complex interactions, ML might be able to help basic research methods raise to new frontiers in a large scale of fields, especially in the setting of high-throughput data like omics. Further, ML approaches might be used to streamline image processing [34] and, therefore, increase and standardize workflows traditionally prone to require high amounts of work force.

In the light of the complex interplay of variables used in ML, surpassing the capabilities of the human brain, it will be more important than ever to validate those algorithms meticulously in prospective clinical trials [3]. After all, even if we may not be able to grasp the individual calculations the ML algorithm makes, we must be able to trust in the answers it provides.

Declarations

Conflict of interest Philipp Lurz: Institutional grants from Abbott Vascular, ReCor and Edwards Lifesciences. Karl-Patrik Kresoja, Matthias Unterhuber, Rolf Wachter and Holger Thiele have nothing to disclose.

References

1. Ahmad F, McNally EM, Ackerman MJ, Baty LC, Day SM, Kullo IJ, Madueme PC, Maron MS, Martinez MW, Salberg L, Taylor MR, Wilcox JE (2019) Establishment of specialized clinical cardiovascular genetics programs: recognizing the need and meeting standards: a scientific statement from the American Heart Association. Circ Genom Precis Med 12:e000054. https://doi.org/10.1161/HCG.0000000000000054

2. Akella A, Akella S (2021) Machine learning algorithms for predicting coronary artery disease: efforts toward an open source solution. Future Sci OA 7:FSO698. https://doi.org/10.2144/fsoa-2020-0206

3. Anderson M, Anderson SL (2019) How should AI be developed, validated, and implemented in patient care? AMA J Ethics 21:E125-130. https://doi.org/10.1001/amajethics.2019.125

4. Antman EM, Loscalzo J (2016) Precision medicine in cardiology. Nat Rev Cardiol 13:591–602. https://doi.org/10.1038/nrcardio.2016.101

5. Asch FM, Poilvert N, Abraham T, Jankowski M, Cleve J, Adams M, Romano N, Hong H, Mor-Avi V, Martin RP, Lang RM (2019) Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert. Circ Cardiovasc

Imaging 12:e009303. https://doi.org/10.1161/CIRCIMAGING.119.009303

6. Atienza F, Martinez-Alzamora N, Velasco JA de, Dreiseitl S, Ohno-Machado L (2000) Risk stratification in heart failure using artificial neural networks. In: Proc AMIA Symp, pp 32–36

7. Badillo S, Banfai B, Birzele F, Davydov II, Hutchinson L, Kam-Thong T, Siebourg-Polster J, Steiert B, Zhang JD (2020) An introduction to machine learning. Clin Pharmacol Ther 107:871–885. https://doi.org/10.1002/cpt.1796

8. Bayés-Genís A, Lanfear D (2019) Liquid biopsy and ehealth in heart failure: the future is now. J Am Coll Cardiol 73:2206–2208. https://doi.org/10.1016/j.jacc.2019.01.071

9. Bazoukis G, Stavrakis S, Zhou J, Bollepalli SC, Tse G, Zhang Q, Singh JP, Armoundas AA (2020) Machine learning versus conventional clinical methods in guiding management of heart failure patients—a systematic review. Heart Fail Rev 26:23–34. https://doi.org/10.1007/s10741-020-10007-3

10. Bojer CS, Meldgaard JP (2021) Kaggle forecasting competitions: an overlooked learning opportunity. Int J Forecast 37:587–603. https://doi.org/10.1016/j.ijforecast.2020.07.007

11. Breiman L (2001) Bagging predictors. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

12. Burghardt TP, Ajtai K (2018) Neural/Bayes network predictor for inheritable cardiac disease pathogenicity and phenotype. J Mol Cell Cardiol 119:19–27. https://doi.org/10.1016/j.yjmcc.2018.04.006

13. Bzdok D, Altman N, Krzywinski M (2018) Statistics versus machine learning. Nat Methods 15:233–234. https://doi.org/10.1038/nmeth.4642

14. de Caestecker M, Humphreys BD, Liu KD, Fissell WH, Cerda J, Nolin TD, Askenazi D, Mour G, Harrell FE, Pullen N, Okusa MD, Faubel S (2015) Bridging translation by improving preclinical study design in AKI. J Am Soc Nephrol 26:2905–2916. https://doi.org/10.1681/ASN.2015070832

15. Ceglarek U, Schellong P, Rosolowski M, Scholz M, Willenberg A, Kratzsch J, Zeymer U, Fuernau G, de Waha-Thiele S, Büttner P, Jobs A, Freund A, Desch S, Feistritzer H-J, Isermann B, Thiery J, Pöss J, Thiele H (2021) The novel cystatin C, lactate, interleukin-6, and N-terminal pro-B-type natriuretic peptide (CLIP)-based mortality risk score in cardiogenic shock after acute myocardial infarction. Eur Heart J 42:2344–2352. https://doi.org/10.1093/eurheartj/ehab110

16. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system, vol 11, pp 785–794. https://doi.org/10.1145/2939672.2939785

17. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP (2020) Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol 9:14. https://doi.org/10.1167/tvst.9.2.14

18. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP (2016) Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med 44:368–374. https://doi.org/10.1097/CCM.0000000000001571

19. Clerx M, Heijman J, Collins P, Volders PGA (2018) Predicting changes to INa from missense mutations in human SCN5A. Sci Rep 8:12797. https://doi.org/10.1038/s41598-018-30577-5

20. Commandeur F, Slomka PJ, Goeller M, Chen X, Cadet S, Razipour A, McElhinney P, Gransar H, Cantu S, Miller RJH, Rozanski A, Achenbach S, Tamarappoo BK, Berman DS, Dey D (2020) Machine learning to predict the long-term risk of myocardial infarction and cardiac death based on clinical risk, coronary calcium, and epicardial adipose tissue: a prospective study. Cardiovasc Res 116:2216–2225. https://doi.org/10.1093/cvr/cvz321

21. Senn S (2005) Dichotomania: an obsessive compulsive disorder that is badly affecting the quality of analysis of pharmaceutical

trials. https://www.isi-web.org/isi.cbs.nl/iamamember/CD6-Sydney2005/ISI2005_Papers/398.pdf

22. Ernst AF, Albers CJ (2017) Regression assumptions in clinical psychology research practice—a systematic review of common misconceptions. PeerJ. https://doi.org/10.7717/peerj.3323

23. Frizzell JD, Liang L, Schulte PJ, Yancy CW, Heidenreich PA, Hernandez AF, Bhatt DL, Fonarow GC, Laskey WK (2017) Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. JAMA Cardiol 2:204–209. https://doi.org/10.1001/jamacardio.2016.3956

24. Georgevici AI, Terblanche M (2019) Neural networks and deep learning: a brief introduction. Intensive Care Med 45:712–714. https://doi.org/10.1007/s00134-019-05537-w

25. Ho JE, Enserro D, Brouwers FP, Kizer JR, Shah SJ, Psaty BM, Bartz TM, Santhanakrishnan R, Lee DS, Chan C, Liu K, Blaha MJ, Hillege HL, van der Harst P, van Gilst WH, Kop WJ, Gansevoort RT, Vasan RS, Gardin JM, Levy D, Gottdiener JS, de Boer RA, Larson MG (2016) Predicting heart failure with preserved and reduced ejection fraction: the international collaboration on heart failure subtypes. Circ Heart Fail. https://doi.org/10.1161/CIRCHEARTFAILURE.115.003116

26. Jaganathan K, KyriazopoulouPanagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, Chow ED, Kanterakis E, Gao H, Kia A, Batzoglou S, Sanders SJ, Farh KK-H (2019) Predicting splicing from primary sequence with deep learning. Cell 176:535-548.e24. https://doi.org/10.1016/j.cell.2018.12.015

27. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, Ashley E, Dudley JT (2018) Artificial intelligence in cardiology. J Am Coll Cardiol 71:2668–2679. https://doi.org/10.1016/j.jacc.2018.03.521

28. Johnston BC, Alonso-Coello P, Friedrich JO, Mustafa RA, Tikkinen KAO, Neumann I, Vandvik PO, Akl EA, Da Costa BR, Adhikari NK, Dalmau GM, Kosunen E, Mustonen J, Crawford MW, Thabane L, Guyatt GH (2016) Do clinicians understand the size of treatment effects? A randomized survey across 8 countries. CMAJ 188:25–32. https://doi.org/10.1503/cmaj.150430

29. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2

30. Kahneman D (2012) Thinking, fast and slow. Penguin psychology. Penguin Books, London

31. Karwath A, Bunting KV, Gill SK, Tica O, Pendleton S, Aziz F, Barsky AD, Chernbumroong S, Duan J, Mobley AR, Cardoso VR, Slater L, Williams JA, Bruce E-J, Wang X, Flather MD, Coats AJS, Gkoutos GV, Kotecha D (2021) Redefining β-blocker response in heart failure patients with sinus rhythm and atrial fibrillation: a machine learning cluster analysis. Lancet 398:1427–1435. https://doi.org/10.1016/S0140-6736(21)01638-X

32. Katz DH, Deo RC, Aguilar FG, Selvaraj S, Martinez EE, Beussink-Nelson L, Kim K-YA, Peng J, Irvin MR, Tiwari H, Rao DC, Arnett DK, Shah SJ (2017) Phenomapping for the identification of hypertensive patients with the myocardial substrate for heart failure with preserved ejection fraction. J Cardiovasc Transl Res 10:275–284. https://doi.org/10.1007/s12265-017-9739-z

33. Kimenai DM, Henry RMA, van der Kallen CJH, Dagnelie PC, Schram MT, Stehouwer CDA, van Suijlen JDE, Niens M, Bekers O, Sep SJS, Schaper NC, van Dieijen-Visser MP, Meex SJR

(2016) Direct comparison of clinical decision limits for cardiac troponin T and I. Heart 102:610–616. https://doi.org/10.1136/heartjnl-2015-308917

34. Kleppe A, Skrede O-J, de Raedt S, Liestøl K, Kerr DJ, Danielsen HE (2021) Designing deep learning studies in cancer diagnostics. Nat Rev Cancer 21:199–211. https://doi.org/10.1038/s41568-020-00327-9

35. Krittanawong C, Johnson KW, Choi E, Kaplin S, Venner E, Murugan M, Wang Z, Glicksberg BS, Amos CI, Schatz MC, Tang WHW (2022) Artificial intelligence and cardiovascular genetics. Life (Basel). https://doi.org/10.3390/life12020279

36. Lahav O, Mastronarde N, van der Schaar M (2018) What is interpretable? Using machine learning to design interpretable decision-support systems. ML4H. https://arxiv.org/abs/1811.10799

37. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539

38. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, Anand I, Maggioni A, Burton P, Sullivan MD, Pitt B, Poole-Wilson PA, Mann DL, Packer M (2006) The Seattle Heart Failure Model: prediction of survival in heart failure. Circulation 113:1424–1433. https://doi.org/10.1161/CIRCULATIONAHA.105.584102

39. Li C, Sun D, Liu J, Li M, Zhang B, Liu Y, Wang Z, Wen S, Zhou J (2019) A prediction model of essential hypertension based on genetic and environmental risk factors in Northern Han Chinese. Int J Med Sci 16:793–799. https://doi.org/10.7150/ijms.33967

40. Li L, Cheng W-Y, Glicksberg BS, Gottesman O, Tamler R, Chen R, Bottinger EP, Dudley JT (2015) Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Sci Transl Med 7:311ra174. https://doi.org/10.1126/scitranslmed.aaa9364

41. Lim N-K, Lee J-Y, Lee J-Y, Park H-Y, Cho M-C (2015) The role of genetic risk score in predicting the risk of hypertension in the Korean population: Korean Genome and Epidemiology Study. PLoS ONE 10:e0131603. https://doi.org/10.1371/journal.pone.0131603

42. Liu L, Chen X, Petinrin OO, Zhang W, Rahaman S, Tang Z-R, Wong K-C (2021) Machine learning protocols in early cancer detection based on liquid biopsy: a survey. Life (Basel). https://doi.org/10.3390/life11070638

43. Lundberg S, Lee S-I (2017) A unified approach to interpreting model predictions. https://arxiv.org/abs/1705.07874

44. Luo Y, Ahmad FS, Shah SJ (2017) Tensor factorization for precision medicine in heart failure with preserved ejection fraction. J Cardiovasc Transl Res 10:305–312. https://doi.org/10.1007/s12265-016-9727-8

45. Malki MA, Pearson ER (2020) Drug-drug-gene interactions and adverse drug reactions. Pharmacogenom J 20:355–366. https://doi.org/10.1038/s41397-019-0122-0

46. Marcu Ş-B, Tăbîrcă S, Tangney M (2022) An overview of Alphafold's breakthrough. Front Artif Intell 5:875587. https://doi.org/10.3389/frai.2022.875587

47. Márquez-Luna C, Loh P-R, Price AL (2017) Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genet Epidemiol 41:811–823. https://doi.org/10.1002/gepi.22083

48. Maulud D, Abdulazeez AM (2020) A review on linear regression comprehensive in machine learning. JASTT 1:140–147. https://doi.org/10.38094/jastt1457

49. Menon S, Damian A, Hu S, Ravi N, Rudin C (2020) PULSE: self-supervised photo upsampling via latent space exploration of generative models. https://arxiv.org/abs/2003.03808

50. Meskó B, Görög M (2020) A short guide for medical professionals in the era of artificial intelligence. NPJ Digit Med 3:126. https://doi.org/10.1038/s41746-020-00333-z

51. Miotto R, Li L, Kidd BA, Dudley JT (2016) Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 6:26094. https://doi.org/10.1038/srep26094

52. Moen E, Bannon D, Kudo T, Graf W, Covert M, van Valen D (2019) Deep learning for cellular image analysis. Nat Methods 16:1233–1246. https://doi.org/10.1038/s41592-019-0403-1

53. Mohammad MA, Olesen KKW, Koul S, Gale CP, Rylance R, Jernberg T, Baron T, Spaak J, James S, Lindahl B, Maeng M, Erlinge D (2022) Development and validation of an artificial neural network algorithm to predict mortality and admission to hospital for heart failure after myocardial infarction: a nationwide population-based study. Lancet Digital Health 4:e37–e45. https://doi.org/10.1016/S2589-7500(21)00228-4

54. Molenaar MA, Selder JL, Nicolas J, Claessen BE, Mehran R, Bescós JO, Schuuring MJ, Bouma BJ, Verouden NJ, Chamuleau SAJ (2022) Current state and future perspectives of artificial intelligence for automated coronary angiography imaging analysis in patients with ischemic heart disease. Curr Cardiol Rep 24:365–376. https://doi.org/10.1007/s11886-022-01655-y

55. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, Andreini D, Budoff MJ, Cademartiri F, Callister TQ, Chang H-J, Chinnaiyan K, Chow BJW, Cury RC, Delago A, Gomez M, Gransar H, Hadamitzky M, Hausleiter J, Hindoyan N, Feuchtner G, Kaufmann PA, Kim Y-J, Leipsic J, Lin FY, Maffei E, Marques H, Pontone G, Raff G, Rubinshtein R, Shaw LJ, Stehli J, Villines TC, Dunning A, Min JK, Slomka PJ (2017) Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. Eur Heart J 38:500–507. https://doi.org/10.1093/eurheartj/ehw188

56. Musoro JZ, Zwinderman AH, Puhan MA, ter Riet G, Geskus RB (2014) Validation of prediction models based on lasso regression with multiply imputed data. BMC Med Res Methodol. https://doi.org/10.1186/1471-2288-14-116

57. Naess H, Logallo N, Waje-Andreassen U, Thomassen L, Kvistad CE (2019) U-shaped relationship between hemoglobin level and severity of ischemic stroke. Acta Neurol Scand 140:56–61. https://doi.org/10.1111/ane.13100

58. Niiranen TJ, Havulinna AS, Langén VL, Salomaa V, Jula AM (2016) Prediction of blood pressure and blood pressure change with a genetic risk score. J Clin Hypertens (Greenwich) 18:181–186. https://doi.org/10.1111/jch.12702

59. Noseworthy PA, Attia ZI, Behnken EM, Giblon RE, Bews KA, Liu S, Gosse TA, Linn ZD, Deng Y, Yin J, Gersh BJ, Graff-Radford J, Rabinstein AA, Siontis KC, Friedman PA, Yao X (2022) Artificial intelligence-guided screening for atrial fibrillation using electrocardiogram during sinus rhythm: a prospective non-randomised interventional trial. Lancet 400:1206–1212. https://doi.org/10.1016/S0140-6736(22)01637-3

60. Oguz C, Sen SK, Davis AR, Fu Y-P, O'Donnell CJ, Gibbons GH (2017) Genotype-driven identification of a molecular network predictive of advanced coronary calcium in ClinSeq® and Framingham Heart Study cohorts. BMC Syst Biol 11:99. https://doi.org/10.1186/s12918-017-0474-5

61. Ortiz J, Ghefter CG, Silva CE, Sabbatini RM (1995) One-year mortality prognosis in heart failure: a neural network approach based on echocardiographic data. J Am Coll Cardiol 26:1586–1593. https://doi.org/10.1016/0735-1097(95)00385-1

62. Pocock SJ, Ariti CA, McMurray JJV, Maggioni A, Køber L, Squire IB, Swedberg K, Dobson J, Poppe KK, Whalley GA, Doughty RN (2013) Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. Eur Heart J 34:1404–1413. https://doi.org/10.1093/eurheartj/ehs337

63. Quer G, Arnaout R, Henne M, Arnaout R (2021) Machine learning and the future of cardiovascular care: JACC state-of-the-art review. J Am Coll Cardiol 77:300–313. https://doi.org/10.1016/j.jacc.2020.11.030

64. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E (2021) Using machine learning approaches for multi-omics data analysis: a review. Biotechnol Adv 49:107739. https://doi.org/10.1016/j.biotechadv.2021.107739

65. Rowe M (2019) An introduction to machine learning for clinicians. Acad Med 94:1433–1436. https://doi.org/10.1097/ACM.0000000000002792

66. Russell SJ, Norvig P (2021) Artificial intelligence: a modern approach, Fourth Edition Pearson series in artificial intelligence. Pearson, Hoboken

67. Samuel AL (1959) Some studies in machine learning using the game of checkers. IBM J Res Dev 3:210–229. https://doi.org/10.1147/rd.33.0210

68. Selkoe DJ (2003) Folding proteins in fatal ways. Nature 426:900–904. https://doi.org/10.1038/nature02264

69. Shah SJ, Katz DH, Deo RC (2014) Phenotypic spectrum of heart failure with preserved ejection fraction. Heart Fail Clin 10:407–418. https://doi.org/10.1016/j.hfc.2014.04.008

70. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, Bonow RO, Huang C-C, Deo RC (2015) Phenomapping for novel classification of heart failure with preserved ejection fraction. Circulation 131:269–279. https://doi.org/10.1161/CIRCULATIONAHA.114.010637

71. Shah SJ, Kitzman DW, Borlaug BA, van Heerebeek L, Zile MR, Kass DA, Paulus WJ (2016) Phenotype-specific treatment of heart failure with preserved ejection fraction: a multiorgan roadmap. Circulation 134:73–90. https://doi.org/10.1161/CIRCULATIONAHA.116.021884

72. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP (2018) Machine learning in cardiovascular medicine: are we there yet? Heart 104:1156–1164. https://doi.org/10.1136/heartjnl-2017-311198

73. Skolnick J, Gao M, Zhou H, Singh S (2021) AlphaFold 2: why it works and its implications for understanding the relationships of protein sequence, structure, and function. J Chem Inf Model 61:4827–4831. https://doi.org/10.1021/acs.jcim.1c01114

74. Sounderajah V, Ashrafian H, Golub RM, Shetty S, de Fauw J, Hooft L, Moons K, Collins G, Moher D, Bossuyt PM, Darzi A, Karthikesalingam A, Denniston AK, Mateen BA, Ting D, Treanor D, King D, Greaves F, Godwin J, Pearson-Stuttard J, Harling L, McInnes M, Rifai N, Tomasev N, Normahani P, Whiting P, Aggarwal R, Vollmer S, Markar SR, Panch T, Liu X (2021) Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open 11:e047709. https://doi.org/10.1136/bmjopen-2020-047709

75. Sousa T, Correia J, Pereira V, Rocha M (2021) Generative deep learning for targeted compound design. J Chem Inf Model 61:5343–5361. https://doi.org/10.1021/acs.jcim.0c01496

76. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25:44–56. https://doi.org/10.1038/s41591-018-0300-7

77. Unterhuber M, Rommel K-P, Kresoja K-P, Lurz J, Kornej J, Hindricks G, Scholz M, Thiele H, Lurz P (2021) Deep learning detects heart failure with preserved ejection fraction using a baseline electrocardiogram. Eur Heart J Digit Health 2:699–703. https://doi.org/10.1093/ehjdh/ztab081

78. Unterhuber M, Kresoja K-P, Rommel K-P, Besler C, Baragetti A, Klöting N, Ceglarek U, Blüher M, Scholz M, Catapano AL, Thiele H, Lurz P (2021) Proteomics-enabled deep learning machine algorithms can enhance prediction of mortality. J Am Coll Cardiol 78:1621–1631. https://doi.org/10.1016/j.jacc.2021.08.018

79. Unterhuber M, Kresoja K-P, Lurz P, Thiele H (2022) Artificial intelligence in proteomics: new frontiers from risk prediction to treatment? Eur Heart J. https://doi.org/10.1093/eurheartj/ehac391

80. Upadhyay AK, Sowdhamini R (2016) Genome-wide prediction and analysis of 3D-domain swapped proteins in the human genome from sequence information. PLoS ONE. https://doi.org/10.1371/journal.pone.0159627

81. Wang X, Luan Y, Yue F (2022) EagleC: a deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps. Sci Adv 8:eabn9215. https://doi.org/10.1126/sciadv.abn9215

82. Williams MN, Grajales CAG, Kurkiewicz D (2013) Assumptions of multiple regression: correcting two misconceptions. University of Massachusetts Amherst

83. Xiong G, Kola D, Heo R, Elmore K, Cho I, Min JK (2015) Myocardial perfusion analysis in cardiac computed tomography angiographic images at rest. Med Image Anal 24:77–89. https://doi.org/10.1016/j.media.2015.05.010

84. Yadav SS, Jadhav SM (2019) Deep convolutional neural network based medical image classification for disease diagnosis. J Big Data. https://doi.org/10.1186/s40537-019-0276-2

85. 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, Deo RC (2018) Fully automated echocardiogram interpretation in clinical practice. Circulation 138:1623–1635. https://doi.org/10.1161/CIRCULATIONAHA.118.034338

86. Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Telenti A (2019) A primer on deep learning in genomics. Nat Genet 51:12–18. https://doi.org/10.1038/s41588-018-0295-5