

Clinical applications of artificial intelligence in cardiology on the verge of the decade

Konrad Pieszko^{1,2}, Jarosław Hiczekiewicz^{1,2}, Jan Budzianowski^{1,2}, Bogdan Musielak^{1,2}, Dariusz Hiczekiewicz^{1,2}, Wojciech Faron¹, Janusz Rzeźniczak³, Paweł Burchardt^{3,4}

¹Clinical Department of Cardiology, Nowa Sol Multidisciplinary Hospital, Nowa Sol, Poland

²University of Zielona Gora, Poland

³Cardiology Clinic, Jozef Strus Hospital, Poznan, Poland

⁴Department of Biology and Environmental Protection, Poznan University of Medical Sciences, Poznan, Poland

Abstract

Artificial intelligence (AI) has been hailed as the fourth industrial revolution and its influence on people's lives is increasing. The research on AI applications in medicine is progressing rapidly. This revolution shows promise for more precise diagnoses, streamlined workflows, increased accessibility to healthcare services and new insights into ever-growing population-wide datasets. While some applications have already found their way into contemporary patient care, we are still in the early days of the AI-era in medicine.

Despite the popularity of these new technologies, many practitioners lack an understanding of AI methods, their benefits, and pitfalls. This review aims to provide information about the general concepts of machine learning (ML) with special focus on the applications of such techniques in cardiovascular medicine. It also sets out the current trends in research related to medical applications of AI.

Along with new possibilities, new threats arise — acknowledging and understanding them is as important as understanding the ML methodology itself. Therefore, attention is also paid to the current opinions and guidelines regarding the validation and safety of AI-powered tools. (Cardiol J 2021; 28, 3: 460–472)

Key words: machine learning, artificial intelligence, cardiology

Introduction

Medical practitioners build their clinical experience when treating thousands of patients during their lifetime. However, nobody lives long enough to experience all possible variants and cases personally. Moreover, the perception and decision making of physicians may vary over time depending on different factors e.g. fatigue, which was reported to affect a physicians' performance in many studies [1]. Constantly dealing with large amounts of data in different modalities is the norm. This is where machines offer their computational advantage

as they can easily digest enormous quantities of data. Machine learning (ML) can be understood as a fundamental technology required to meaningfully process data that exceeds the capacity and comprehensive abilities of a human brain [2].

Artificial intelligence (AI) is often described as software allowing computer systems to perform tasks that are believed to require human intelligence. This is an umbrella term for many computational methods, some of which are recently attracting a lot of attention from the medical community. The advantages of a computerized approach over medical data analysis include lowered cost,

Address for correspondence: Konrad Pieszko, MD, Clinical Department of Cardiology, Nowa Sol Multidisciplinary Hospital, ul. Chałubińskiego 7, 67–100 Nowa Sól, Poland, tel: +48 728 358 082, e-mail: konrad.pieszko@gmail.com

Received: 18.02.2020

Accepted: 25.05.2020

This article is available in open access under Creative Commons Attribution-Non-Commercial-No Derivatives 4.0 International (CC BY-NC-ND 4.0) license, allowing to download articles and share them with others as long as they credit the authors and the publisher, but without permission to change them in any way or use them commercially.

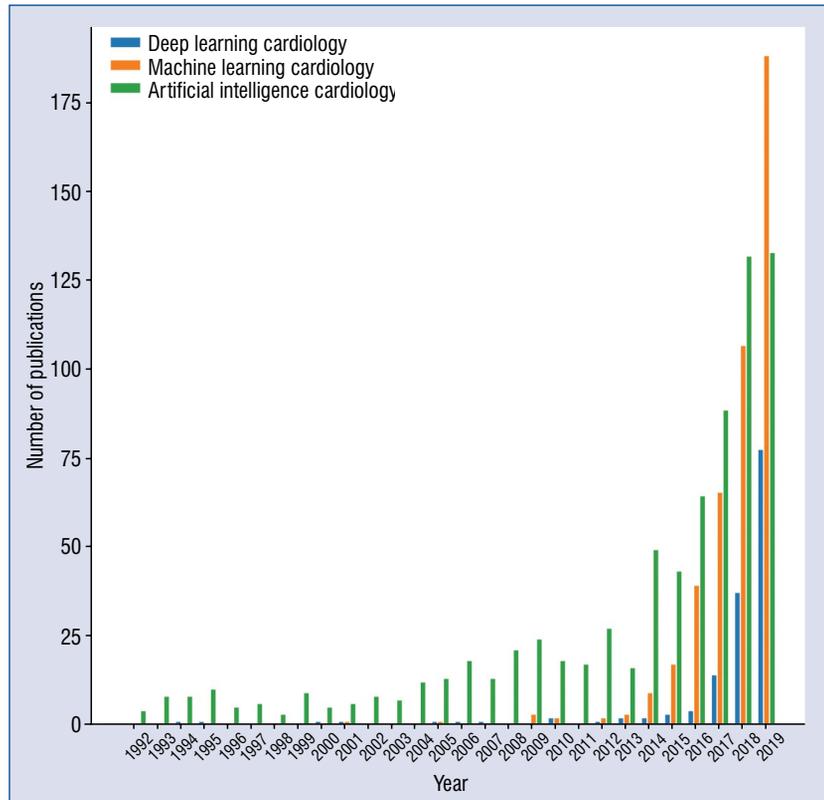


Figure 1. The number of publications related to artificial medicine over the last 15 years. Each series represents the number of results found in PubMed matching the given phrase by the year of publication.

increased speed and accessibility. In this review, some of the most prominent and promising practical applications of ML techniques in the field of cardiology are described. Also discussed are the potential safety issues related to the use of AI in clinical practice.

Recent years have triggered a rapid growth in ML-related publications in all fields of medicine. The growth in interest in this area has been exponential as illustrated in Figure 1. Although the greatest progress in the field of AI happened over the last 10 years, the onset of AI can be traced back much earlier. The historical aspect of AI is discussed in more detail by Benjamins et al. [3].

Owing to the popularity of this topic, many reviews have been published with the aim of familiarizing the reader with the relatively new concepts in AI methods [2, 4]. The specific applications of AI in cardiology have also been reviewed [3, 5–8]. Some of these papers focused on general usage scenarios of ML in cardiology [9] while others provided deeper insights into specific applications e.g. image analysis [10–12]. Some reviews covered technical aspects of various ML methods in greater detail [5]. Although many reviews have

already been published, the field of medical AI is progressing very rapidly, and new research is published almost every day. This paper aims to present some of the most recent applications of AI in cardiology and discusses many safety concerns, which have recently received a lot of attention from the scientific community.

Overview of artificial intelligence and machine learning

The most commonly used terms “artificial intelligence” (AI) and “machine learning” (ML) are interrelated and are sometimes used in a similar context. However, they do require some disambiguation. Figure 2 presents how the most common techniques relate to each other.

Artificial intelligence is often described as human-like intelligence demonstrated by a machine. This is a broad term that applies to systems based on ML as well as to expert systems and robotics. ML, by contrast, is a group of algorithms that allow a computer to learn to perform a specific task based on several examples.

Machine learning algorithms are rooted in so-called traditional statistics. The simplest ML model

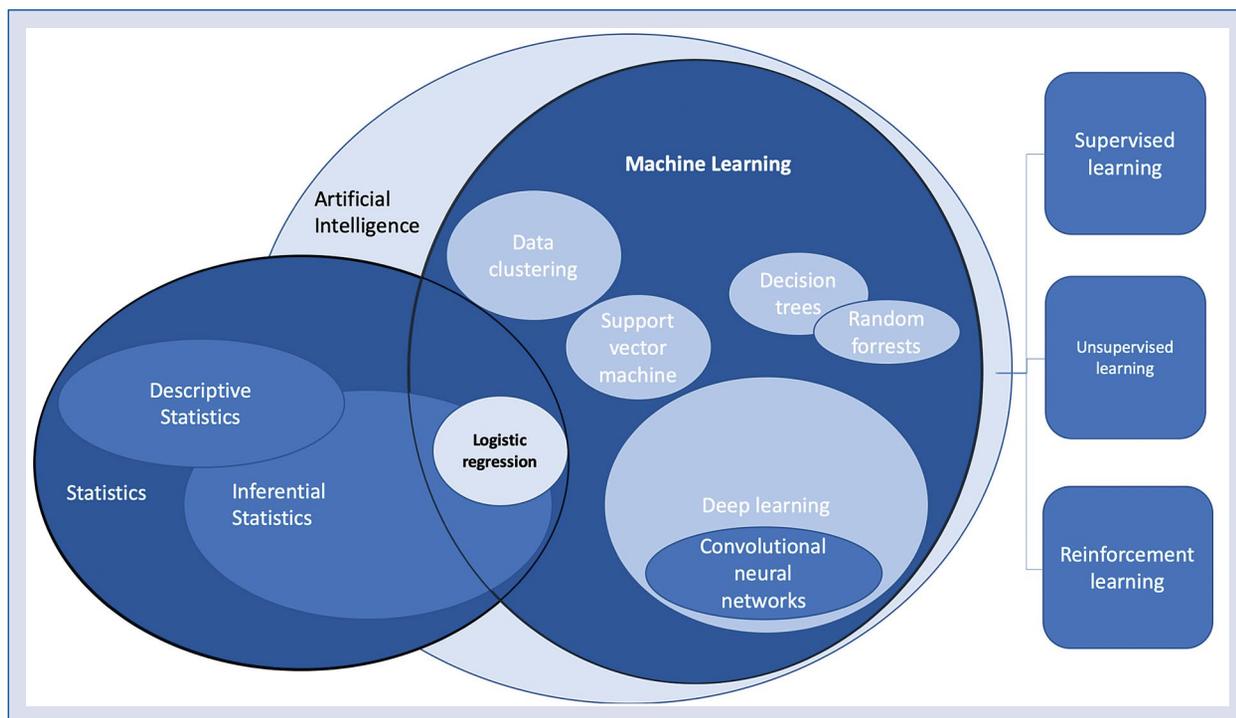


Figure 2. The machine learning and statistical methods and their relation to each other. On the right: three most common machine learning paradigms.

can be based on logistic regression. However, more sophisticated methods including decision trees, support vector machines, random forests or neural networks that have the advantage of handling complex and non-linear relationships within the data while avoiding ‘improper dichotomization’ [5]. The most recently developed techniques include deep neural networks, also called deep learning. These algorithms allow for the rapid progress of image recognition, natural language processing, speech recognition and are widely used in the latest medical research [2, 13].

Machine learning workflow

There are three standard ways in which any machine-learning model can be trained: supervised, unsupervised and reinforced (so called reinforcement learning). The first paradigm takes advantage of a set of labeled data — examples of input data along with correct answers. The dataset is divided into training, validation and test sets at an early stage of data manipulation. The training set is used to create the model — this is the data the algorithm learns from. To assess the process of learning, the validation set is used. During the process of training, the performance of the model is assessed multiple times on the validation set and the model improves

gradually. In some cases, samples from the training and validation sets are shuffled in a process called cross-validation. The test set is used when the training process is finished to assess model performance on unseen data. The choice of sizes of these sets is based on available resources and depends on several factors. Generally, the more cases there are in the training set, the better the model performs. However, at some point, the model performance reaches a plateau and does not improve significantly despite adding new cases. In a cornerstone study by Gulshan et al. [14], such a plateau was observed when using 60,000 or more training images. On the other hand, the more cases that are held out in the test/ validation sets, the narrower are the confidence intervals of a classifiers’ performance measures. When comparing two classifiers, the absolute number of cases needed in the test set can be estimated using statistical test power calculations [15]. Figure 3 illustrates the typical workflow for the application of AI in a prediction task.

Interestingly, through unsupervised learning, it is possible to find patterns in the data without explicitly specifying what we are looking for. Various algorithms including hierarchical clustering, k-means clustering, neural networks, and many others can allow for the self-organization of the

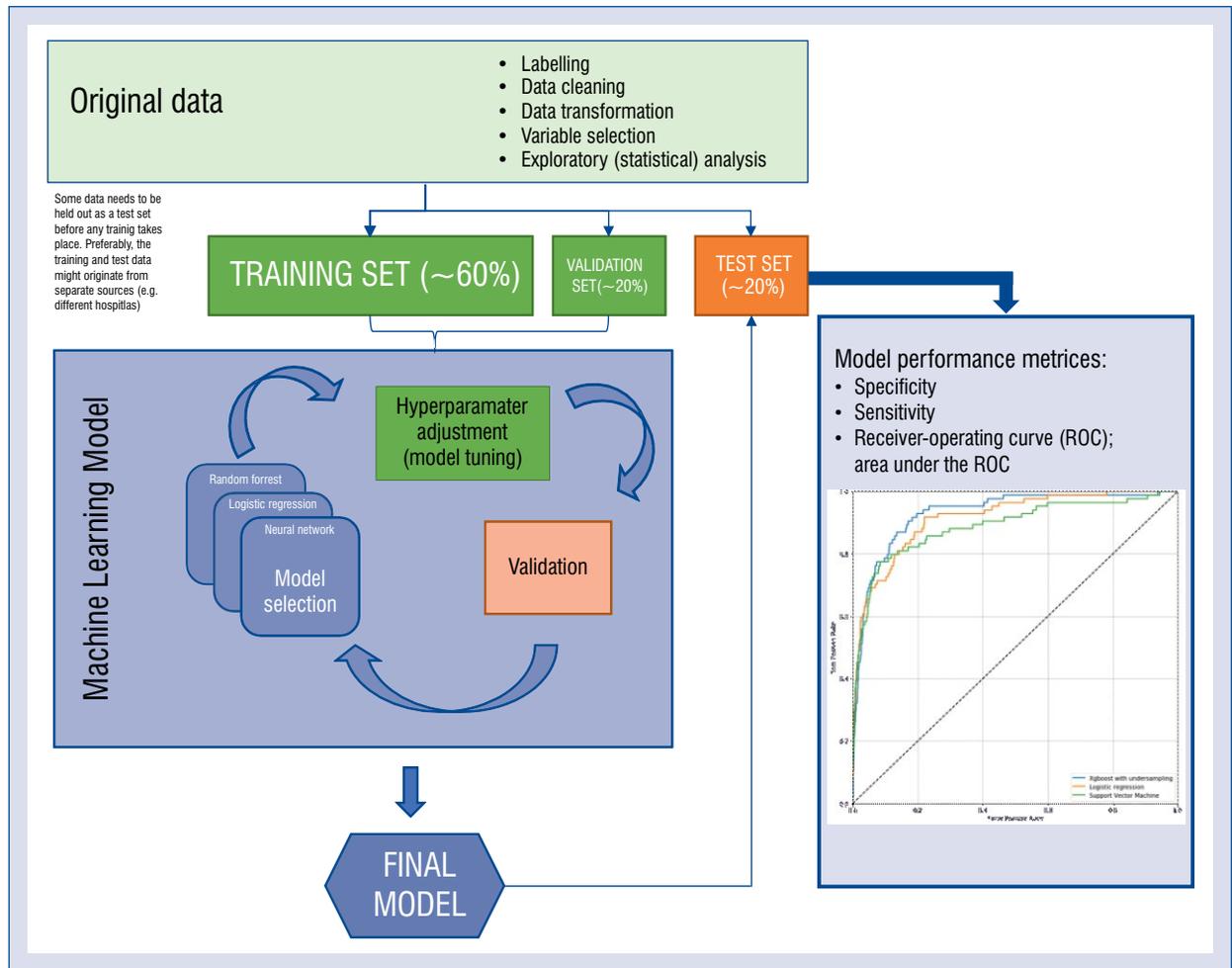


Figure 3. The workflow for a typical task for medical artificial intelligence — supervised learning.

data. This is usually a starting point for analyses using conventional statistical methods. Unsupervised learning imitates human intelligence and the ability to draw conclusions based on the data alone.

In the last paradigm, reinforcement learning, the learning process is continuous — the system works and learns from its own mistakes. These kinds of models are very successful in various applications (a program based on reinforcement learning [AlphaGo] beat the world champion in the game of ‘Go’ in 2017 — a task believed to be impossible for a machine) [16]. Despite being a great area of research, reinforcement learning has not been widely adopted in medical applications yet. As the model can modify its behavior over time, it has not been well established on how to ensure its clinical safety.

Deep learning

Deep neural networks, often referred to as deep learning allowed for a great leap forward in image

recognition and natural language processing including automatic translation, voice recognition, and many other breakthroughs [17]. The image recognition techniques proved to be very useful in the analysis of medical images and physiological signals like electrocardiogram (ECG). The core concept lies in mimicking the way the human visual cortex works.

Bizopoulos and Koutsouris [18] in their systematic review provided a detailed overview of the applications of deep learning in the field of cardiology with insights into the various methodologies and architectures used. They also made a listing of publicly available datasets that can be used for developing and benchmarking ML models.

In this review, working principles and detailed properties of various ML algorithms are not discussed. Anyone interested in acquiring more detailed knowledge can refer to the article by Johnson et al. [5] where they describe the working principles of a number of ML methods along with their advantages and disadvantages.

How can artificial intelligence contribute to the area of cardiology?

In many usage scenarios AI is designed to mimic actions typically performed by a doctor — it recognizes a disease in images or classifies some other signals to provide an answer that a trained specialist would base on the same data. However, one of the most inspiring applications of AI is where it can provide insights into the data that were not anticipated by finding patterns in high-dimensional data. A canonical example is the study of diabetic retinopathy images — the algorithm was designed to determine a patient's cardiovascular risk based on an eye fundus image alone [19]. It turned out that the model was not only able to predict the total risk but also could determine the individual risk factors (sex, age, systolic blood pressure) to a degree of precision not reported before. Such results not only have a great potential for practical application but can also guide basic research given that sex differences in eye fundus are not yet fully understood.

Another example of an ML model analyzes data in a different way than doctors are used to doing was presented by researchers from the Mayo Clinic [20]. Using a deep neural network trained on over 600,000 ECGs, they showed that it was possible to identify the 'electrocardiographic signature' of paroxysmal atrial fibrillation in a standard 12-lead 10-second ECG taken during sinus rhythm. They achieved an area under the receiver-operating curve of 0.9, proving this method could be potentially useful for population-wide atrial fibrillation screening.

Soon the same team went even further and developed a model able to predict the presence of asymptomatic left ventricular dysfunction using the digital data from resting ECGs and trans-thoracic echocardiograms of 97,827 subjects [21]. Then, they validated it prospectively [22]. Currently, they are running a randomized clinical trial to investigate the usefulness of the proposed approach for screening for asymptomatic left ventricular dysfunction in primary care settings [23].

Unsupervised learning can be of great value when it comes to discovering new patterns and structures in data. In this technique, data is fed to an algorithm without labels and then becomes self-organized into multiple subgroups based on similarities between data points. This allows for the identification of new, unknown features and can drive further research. Shah et al. [24] used this approach to prospectively study patients with diagnosed heart failure with preserved ejection

fraction. By applying an unsupervised ML method called hierarchical cluster analysis, they were able to classify the subjects into three distinctive phenogroups that differed significantly in terms of outcomes. Another study [25] used a similar methodology to identify groups with a potential substrate for heart failure among hypertensive patients. These are good examples of how unsupervised learning techniques can provide a starting point for further analyses using supervised methods and inferential statistics. Such an approach may lead to the development of more specific treatment strategies according to the paradigm of precision medicine [26]. Very recently, Wang et al. [27] used an unsupervised autoencoder based on a deep learning algorithm to represent data from electronic health records and compared the classification based on these learned features with more conventional approaches.

Table 1 presents a non-exhaustive list of AI research examples in the field of cardiology [20–22, 28–56]. This list is intended to give a general overview of possible areas of application and demonstrates how AI can improve various aspects of patient care in cardiology. The studies are grouped by the type of input data used (imaging data, ECG signal, clinical data). A brief description of the methodology is provided for each example.

AI — hype or hope?

Artificial intelligence has brought as much hope as fear even long before it became a reality. It is even argued that there is little evidence that it improves patient outcomes [57]. This section aims to illustrate some of the potential threats and difficulties that need to be overcome to ensure that medical AI benefits us all.

AI safety

What if a machine makes the wrong diagnosis? There are indeed safety concerns regarding the use of automated decision support systems. One of the issues that can affect the practical safety of an AI-based classification tool is the 'hidden stratification' of the dataset. This term was coined by researchers from the University of Adelaide and Stanford University [58] and describes a situation in supervised learning when, due to the coarse labeling of the data (for example normal vs. abnormal), there are some unrecognized subgroups within each label. Obviously, the machine cannot learn a class if it was not labeled specifically. The system can learn to recognize the more general label quite well while underperforming on some

Table 1. Examples of artificial intelligence applications in cardiology.

Diagnostic modality/type of data used	Application	Study methodology	Reference(s)
Echocardiography	Identification of echocardiographic views	A convolutional neural network was used to distinguish between 15 standard echocardiographic views with an accuracy of 97.8%	Madani et al. 2018 [28]
	Differentiating CP from RCM	The model was based on an associative memory classifier algorithm. Echocardiograms of 50 patients with CP, 44 with RCM and 47 controls were used to train the model	Sengupta et al. 2016 [29]
	Fully automated echocardiogram interpretation and detection of selected clinical conditions	A convolutional neural network was trained on 14,035 echocardiograms to identify views, perform the segmentation of heart chambers, determine ejection fraction and other measurements and finally to detect a number of clinical conditions (cardiomyopathy, cardiac amyloidosis and pulmonary arterial hypertension with the C statistics of 0.93, 0.87, and 0.85, respectively)	Zhang et al. 2018 [30]
CT	Calculating CS based on CT-angiography scans. (May obviate the need for a separate CS scan; thus, reducing the radiation dose)	The authors designed a convolutional neural network that processes each of the three axes (axial, sagittal, coronal) separately. The model was trained using a total of 250 hand-annotated exams	Wolterink et al. 2016 [31]
	Calculating FFR values based on cardiac CT	The models created using convolutional neural networks have some advantages (including shorter computation times) over the clinically validated approach based on computational fluid dynamics while maintaining a non-inferior performance	Coenen et al. 2018 [32] Teschke et al. 2018 [33]
	Predicting all-cause mortality based on cardiac CT and clinical variables	25 clinical and 44 CT-derived variables of over 10,000 patients were used to train the iterative Logit Boost algorithm. The resulting model could predict a 5-year mortality rate with the c-statistic of 0.79	Motwani et al. 2017 [34]
	CT scan denoising — improving readability of acquired images while also reducing the necessary radiation exposure	The authors obtained scans using 20% and 100% of the clinical radiation dose. The model based on generative adversarial network architecture was trained to generate full-quality images based on the images acquired with a low radiation dose	Wolterink et al. 2017 [35]
	Detecting significant coronary lesions based on the motion of the LV myocardium	The complex model consisted of a convolutional neural network (for the myocardium segmentation), an unsupervised convolutional autoencoder (for the extraction of the myocardium characteristics) and a support vector classifier	Zreik et al. 2018 [36]
	Predicting cardiac death after myocardial perfusion SPECT imaging	A total of 122 features (both the clinical data and variables derived from SPECT scans) of over 8,000 patients were used to train the multiple ML models. A model based on SVM outperformed baseline logistic regression as well as random forests	Haro Alonso et al. 2019 [37]
	Detecting the presence and location of significant coronary artery stenosis based on SPECT images	In these multicenter studies, all patients underwent myocardial perfusion imaging and coronary angiography within 6 months. A deep neural network was trained to predict obstructive coronary disease based on SPECT myocardial perfusion images	Betancur et al. 2018, 2019 [38, 39]

→

Table 1 (cont.). Examples of artificial intelligence applications in cardiology.

Diagnostic modality/type of data used	Application	Study methodology	Reference(s)
MRI	Predicting MACE using a combination of clinical data and myocardial perfusion SPECT images	28 clinical variables, 17 stress test variables, and 25 imaging variables of 2,619 patients were analyzed. The ML model was based on the Logit Boost algorithm	Betancur et al. 2018 [40]
	Segmentation of heart structures, automatic measurement of the LV end-diastolic volume and other values	A fully convolutional neural network was trained using pixel-annotated MRI images from 4,875 patients. The model was able to perform highly accurate automatic measurements and delineation of heart structures	Bai et al. 2018 [41]
	Detecting abnormalities of aortic valve	The authors developed a novel strategy for training medical ML models using unlabeled imaging data. They created a weakly-supervised model capable of diagnosing aortic valve abnormalities in MRI scans	Fries et al. 2019 [42]
	Objective assessment of atrial scarring for patients with AF	The authors developed a complete pipeline for atrial scarring segmentation. A classification algorithm based on SVM was used	Yang et al. 2018 [43]
	Diagnosing pulmonary hypertension based on cardiovascular MRI	The model was trained using 220 MRI scans of patients who had also underwent right heart catheterization	Swift et al. 2020 [44]
Coronary angiography	Segmentation of coronary vessels from angiograms	The model was based on a U-Net architecture (a type of a deep neural network). 3,302 still images of coronary arteries were used to train the model	Yang et al. 2019 [45]
ECG signal	Diagnosing ALVD using ECG only	The ECG signals and echocardiographic data of 97,829 patients were used (the time between ECG and echocardiography was less than 2 weeks). A model based on a neural network could predict ALVD with a sensitivity and specificity of 86%. The initial study laid the groundwork for a prospective evaluation and the ongoing clinical trial	Attia et al. 2019 [21,22]
	Detecting paroxysmal AF based on contemporary 12-lead ECG taken on SR	The authors have shown that it is possible to identify an 'electrocardiographic signature' of paroxysmal AF in a routine 10-second 12-lead ECG. The use of a convolutional neural network allowed the detection of signals invisible to the human eye	Attia et al. 2019 [20]
EHR	Predicting the development of moderate to severe MR based on 12-lead ECG using a deep neural network	The AUROC in external validation of 10,865 cases was 0.877. Positively diagnosed patients also had a higher chance of developing MR in the future. Additionally, the authors used visualization techniques that helped understand which parts of an ECG influence the decisions of their algorithm	Kwon et al. 2020 [46]
	Predicting cardiovascular risk based on records from primary care	30 variables identified within the primary health records of 378,256 patients were analyzed. The authors used a number of ML algorithms including logistic regression, random forests and neural networks	Weng et al. 2017 [47]



Table 1 (cont.). Examples of artificial intelligence applications in cardiology.

Diagnostic modality/type of data used	Application	Study methodology	Reference(s)
Clinical data	Predicting the in-hospital mortality rate, re-admission and a prolonged length of stay based on raw electronic health records	Multi-year medical histories stored in EHRs linked to 216,221 hospitalizations were converted into over 46 billion data points, each representing a result, clinical event, physician's note etc. An ensemble of three types of neural networks was trained to predict various clinical endpoints with high accuracy	Rajkumar et al. 2018 [48]
	Predicting the probability of in-hospital death at the time of admission	The model was crated based on retrospective data but validated prospectively and externally in 3 different hospitals. A total number of over 75,000 admissions were used to create and validate the model. The AUROC was 0.86 in an external validation	Brajer et al. 2020 [49]
	Predicting readmission of patients with heart failure	An EHR-wide feature selection (over 4,000 variables were considered) and a model based on logistic regression was developed to predict the 30-day readmission rates	Shameer et al. 2017 [50]
	Predicting long- and short-term mortality after ACS	In these papers various 'classical' ML models (support vector machines, random forests, xgboost) were developed to predict mortality after acute coronary syndromes using clinical data	Shouval et al. 2017 [51] Wallert et al. 2017 [52] Pieszko et al. 2018, 2019 [53, 54]
	Predicting the risk of MACE and bleeding after ACS	The data on over 24,000 patients with ACSs were pooled from 4 randomized controlled trials. The ML algorithm demonstrated superiority over traditional risk scores	Gibson et al. 2020 [55]
	Selecting the right patients for CRT	Classical ML algorithms were applied to predict survival after CRT implantation. The model based on random forest showed the best performance	Kalscheur et al. 2018 [56]

ACS — acute coronary syndrome; AF — atrial fibrillation; ALVD — asymptomatic left ventricular dysfunction; CP — constrictive pericarditis; CRT — cardiac resynchronization therapy; CS — Calcium Score; CT — computed tomography; EHR — electronic health records; ECG — electrocardiogram; FFR — fractional flow reserve; LV — left ventricle; MACE — major adverse cardiac events; ML — machine learning; MR — mitral regurgitation; MRI — magnetic resonance imaging; RCM — restrictive cardiomyopathy; SR — sinus rhythm; SVM — support vector machines; SPECT — single-photon emission computed tomography

specific subtype that was not given a separate label despite different clinical characteristics (Fig. 4). As a result, the reported performance measures can be good, albeit do not reflect the actual clinical usefulness of the model. If the 'hidden' subtype is a more dangerous one, it is clear that the consequences could be serious. This leads to a situation contrary to the common sense of a doctor, who intuitively tries to exclude the most dangerous diagnoses first (even if they are not common). Simply phrased, AI trained by means of an improperly labeled dataset may seem to make few mistakes but may still fail in very important cases, while a doctor could still make mistakes in less important cases. The difference between a computer and a human being lies in the fact that humans understand the consequences

of their decisions and try to do their best when they know that the stakes are high.

Oakden-Rayner et al. [58] recognize the underlying mechanisms that cause these types of errors and they propose several methods to address this issue. The main reason why hidden stratification can occur is the improper labeling of the data (oversimplified labels). One method proposed to prevent the hidden stratification is the use of exhaustive prospective data labeling in a tree-like fashion that includes classes and finer subclasses, which may be additionally weighted given their clinical significance. Such a predefined schema could even be standardized by an external authority and serve for benchmarking the models designed for a similar task. One of the studies that

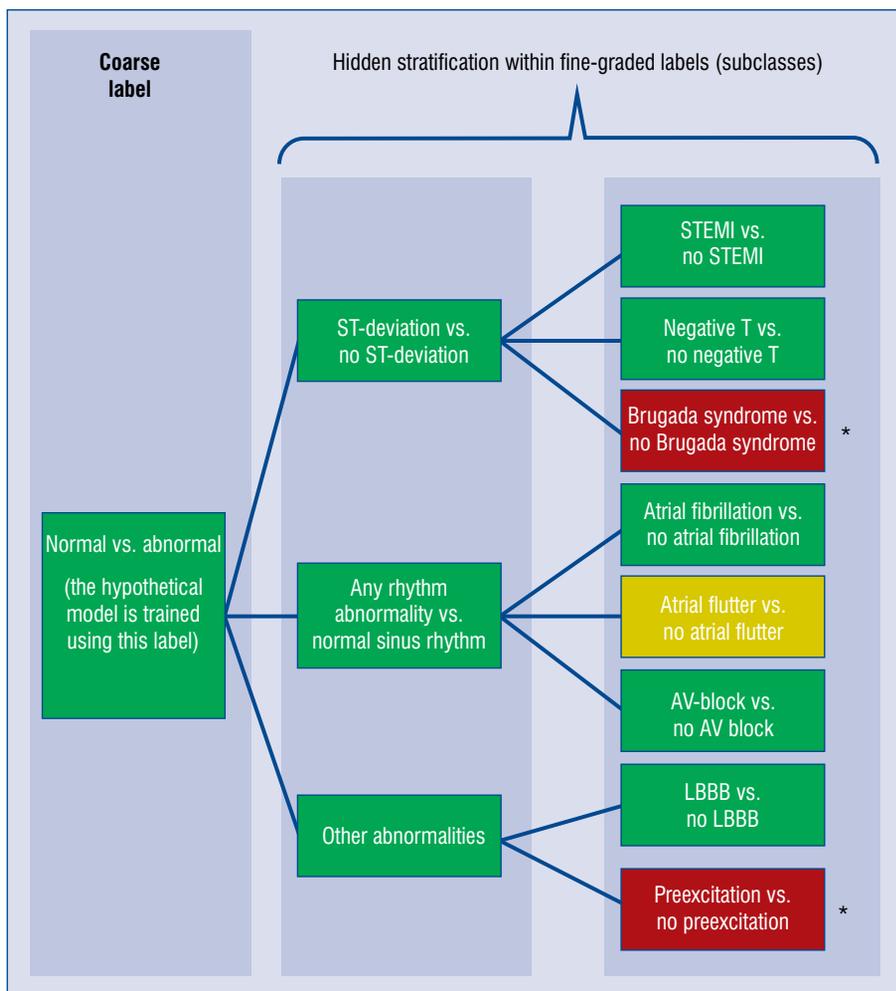


Figure 4. Hidden stratification explained using a hypothetical model for electrocardiogram (ECG) abnormality prediction. The model is trained to detect any abnormality in ECG and it works very well, having high specificity and sensitivity. However, the errors are not equally distributed across possible abnormal conditions. Some of the errors are detected very well while others are ignored by the algorithm. Such cases are rare so the fact that the model does not recognize these abnormal conditions does not significantly affect the overall performance in detecting any ECG abnormality; AV — atrioventricular; LBBB — left bundle branch block; NSTEMI — non-ST-segment elevation myocardial infarction; STEMI — ST-segment elevation myocardial infarction; *rare conditions; predictive performance: green background of a box means good; yellow — moderate; red — poor.

used such a well-defined label structure, including coarse classes and more fine-graded subclasses, was used in the classification of skin lesion images [59].

Explainable AI

Although often perceived as a black-box, there are various methods to provide insights into how ML methods generate their predictions. In contrast to traditional, regression-based statistical inference, more complex methods can model real-world relationships in a better way. In such cases, ML methods can even be used as a source of analytical

insights into the data by providing information on how each variable affects the outcome.

The explanatory features of AI also have their legal aspects. The General Data Protection Regulation, which was introduced in the European Union in 2016, imposes constraints on how personal data can be used in automated decision systems and so-called ‘profiling’, including healthcare applications. These regulations form additional safety measures for the protection of privacy while also guaranteeing the ‘right to explanation’ [60]. They introduce several practical challenges for the design and practical deployment of machine-learning

algorithms. Providing explanatory features (i.e. a way in which the model can communicate what led to its final conclusion) can be seen as an additional safety measure. Such an explanation can be provided in a form of a graph illustrating the importance of features, a heatmap (in the case of images) or a full written report, similar to the one that a physician would write [61]. A very recent study proved that explanatory features increases a doctor's trust as a decision support tool and a willingness to follow its advice. However, there is no evidence that presenting a prediction alongside such an explanation leads to a better clinical outcome than showing the prediction alone. In fact, automation bias could theoretically even lead to worse clinical outcomes [62].

Is the AI biased?

As machine-learned models are finding their way into routine medical practice, concerns are raised regarding the potential statistical bias (that might be introduced into the model) as well as the automation bias (related to the use of AI tools by doctors).

The source of statistical bias lies in the data that was used to train the algorithm, either because it was not representative of the population or because it contained non-ignorable (not randomly distributed) missing values or additional data points (e.g. given that doctors are more likely to perform additional tests if the patient is in poor condition) [63–65]. This kind of bias is nothing new and has always been a concern for prediction models. The use of AI decision support tools can also introduce so-called 'automation bias' which relates to the behavioral aspects seen in automation systems used by humans. This topic was discussed in great detail in the paper by Parasuraman and Manzey [62].

Automation bias is caused by a natural human tendency to pay less attention to automated tasks when under pressure. This happens because the users tend to ascribe greater power and authority to automated aids than to other sources of advice. In other words, a 'human in the loop' is likely to follow the advice of AI, even if other available sources and his own knowledge contradict that.

Clinical trials — Are we there yet?

Despite the rapid development and growing interest in the applications of ML methods in medicine, the majority of up-to-date publications are based on experiments in laboratory settings. There have been very few studies which indicate that using AI-based tools improves patient outcomes.

The fact that a new drug works as expected in pre-clinical experiments does not prove its usefulness and safety. Similarly, studies have shown that achieving good results when testing a ML model in a controlled environment does not necessarily mean that it will improve patient outcomes [66] when used in standard practice. Various psychological factors affect a doctors' response to the suggestions of computer systems and it is known that the presence of such a system can sometimes decrease their vigilance resulting in lower sensitivity. This can be well illustrated by the adoption of computer-aided detection for mammography — a sort of AI algorithm developed in the 1990s in the United States, aimed to assist radiologists in assessing mammograms. The algorithm had been developed before the 'deep learning era' but it seemed to improve breast cancer diagnostics in laboratory settings. It was reimbursed by medical insurers who decided to pay more for assessing radiograms using computer-aided detection. However, in clinical setting it did not only fail to improve radiologists' performance but also decreased their sensitivity [67].

Up till now there have been very few clinical trials related to the use of AI-tools in medicine [68]. One such study proved that AI-assisted colonoscopy could help detect more malignant lesions but it also increased the number of unnecessary biopsies [69]. Another example is the EAGLE trial that was mentioned earlier [23]. A recent randomized control trial evaluated the application of an early warning system against hypotension during elective surgery [70]. Despite many studies conducted prospectively and on a large scale, we still know very little about how the actual application of AI in healthcare affects patients and doctors. Well-designed randomized clinical trials are needed to prove the safety and usefulness of medical AI in real-world settings.

Summary

Artificial intelligence is anticipated to shape the new decade and bring meaningful changes to society, the economy, healthcare and people's lifestyles. For these reasons, AI has been hailed as the fourth industrial revolution [71]. The 'technology of the future' is already here but converting this into an actual benefit for patients is a task that lies with clinicians.

Recent years have seen numerous studies that took advantage of various breeds of AI. In many cases algorithms were designed and

validated using the retrospective data only. We are now entering the phase in which ML models need to be tested prospectively and in clinical settings. Knowing how data-hungry ML models are, it is important to develop and adopt the standards of data acquisition that make it easier to cooperate on multi-center projects. On the other hand, we also need to standardize the tools used to monitor the performance of AI-based systems.

Acknowledgements

The authors would like to thank Dr. Paweł Budzianowski, PhD, for his helpful advice on the various technical issues described in this paper.

Conflict of interest: None declared

References

1. Gates M, Wingert A, Featherstone R, et al. Impact of fatigue and insufficient sleep on physician and patient outcomes: a systematic review. *BMJ Open*. 2018; 8(9): e021967, doi: [10.1136/bmjopen-2018-021967](https://doi.org/10.1136/bmjopen-2018-021967), indexed in Pubmed: 30244211.
2. Rajkomar A, Dean J, Kohane I, et al. Machine Learning in Medicine. *N Engl J Med*. 2019; 380(14): 1347–1358, doi: [10.1056/NEJMr1814259](https://doi.org/10.1056/NEJMr1814259), indexed in Pubmed: 30943338.
3. Benjamins JW, Hendriks T, Knuuti J, et al. A primer in artificial intelligence in cardiovascular medicine. *Neth Heart J*. 2019; 27(9): 392–402, doi: [10.1007/s12471-019-1286-6](https://doi.org/10.1007/s12471-019-1286-6), indexed in Pubmed: 31111458.
4. Liu Y, Chen PHC, Krause J, et al. How to read articles that use machine learning: users' guides to the medical literature. *JAMA*. 2019; 322(18): 1806–1816, doi: [10.1001/jama.2019.16489](https://doi.org/10.1001/jama.2019.16489), indexed in Pubmed: 31714992.
5. Johnson KW, Torres Soto J, Glicksberg BS, et al. Artificial Intelligence in Cardiology. *J Am Coll Cardiol*. 2018; 71(23): 2668–2679, doi: [10.1016/j.jacc.2018.03.521](https://doi.org/10.1016/j.jacc.2018.03.521), indexed in Pubmed: 29880128.
6. Shameer K, Johnson KW, Glicksberg BS, et al. Machine learning in cardiovascular medicine: are we there yet? *Heart*. 2018; 104(14): 1156–1164, doi: [10.1136/heartjnl-2017-311198](https://doi.org/10.1136/heartjnl-2017-311198), indexed in Pubmed: 29352006.
7. Mazzanti M, Shirka E, Gjergo H, et al. Imaging, health record, and artificial intelligence: hype or hope? *Curr Cardiol Rep*. 2018; 20(6): 48, doi: [10.1007/s11886-018-0990-y](https://doi.org/10.1007/s11886-018-0990-y), indexed in Pubmed: 29749590.
8. Seetharam K, Shrestha S, Sengupta PP. Artificial Intelligence in Cardiovascular Medicine. *Curr Treat Options Cardiovasc Med*. 2019; 21(6): 25, doi: [10.1007/s11936-019-0728-1](https://doi.org/10.1007/s11936-019-0728-1), indexed in Pubmed: 31089906.
9. Dorado-Díaz PI, Sampedro-Gómez J, Vicente-Palacios V, et al. Applications of artificial intelligence in cardiology. The future is already here. *Rev Esp Cardiol (Engl Ed)*. 2019; 72(12): 1065–1075, doi: [10.1016/j.rec.2019.05.014](https://doi.org/10.1016/j.rec.2019.05.014), indexed in Pubmed: 31611150.
10. Dey D, Slomka PJ, Leeson P, et al. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *J Am Coll Cardiol*. 2019; 73(11): 1317–1335, doi: [10.1016/j.jacc.2018.12.054](https://doi.org/10.1016/j.jacc.2018.12.054), indexed in Pubmed: 30898208.
11. Siegersma KR, Leiner T, Chew DP, et al. Artificial intelligence in cardiovascular imaging: state of the art and implications for the

- imaging cardiologist. *Neth Heart J*. 2019; 27(9): 403–413, doi: [10.1007/s12471-019-01311-1](https://doi.org/10.1007/s12471-019-01311-1), indexed in Pubmed: 31399886.
12. Litjens G, Ciampi F, Wolterink JM, et al. State-of-the-Art Deep Learning in Cardiovascular Image Analysis. *JACC Cardiovasc Imaging*. 2019; 12(8 Pt 1): 1549–1565, doi: [10.1016/j.jcmg.2019.06.009](https://doi.org/10.1016/j.jcmg.2019.06.009), indexed in Pubmed: 31395244.
13. Litjens G, Ciampi F, Wolterink JM, et al. State-of-the-Art Deep Learning in Cardiovascular Image Analysis. *JACC Cardiovasc Imaging*. 2019; 12(8 Pt 1): 1549–1565, doi: [10.1016/j.jcmg.2019.06.009](https://doi.org/10.1016/j.jcmg.2019.06.009), indexed in Pubmed: 31395244.
14. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016; 316(22): 2402–2410, doi: [10.1001/jama.2016.17216](https://doi.org/10.1001/jama.2016.17216), indexed in Pubmed: 27898976.
15. Beleites C, Neugebauer U, Bocklitz T, et al. Sample size planning for classification models. *Anal Chim Acta*. 2013; 760: 25–33, doi: [10.1016/j.aca.2012.11.007](https://doi.org/10.1016/j.aca.2012.11.007), indexed in Pubmed: 23265730.
16. Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*. 2016; 529(7587): 484–489, doi: [10.1038/nature16961](https://doi.org/10.1038/nature16961), indexed in Pubmed: 26819042.
17. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015; 521(7553): 436–444, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539), indexed in Pubmed: 26017442.
18. Bizopoulos P, Koutsouris D. Deep learning in cardiology. *IEEE Rev Biomed Eng*. 2019; 12: 168–193, doi: [10.1109/RBME.2018.2885714](https://doi.org/10.1109/RBME.2018.2885714), indexed in Pubmed: 30530339.
19. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018; 2(3): 158–164, doi: [10.1038/s41551-018-0195-0](https://doi.org/10.1038/s41551-018-0195-0), indexed in Pubmed: 31015713.
20. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet*. 2019; 394(10201): 861–867, doi: [10.1016/S0140-6736\(19\)31721-0](https://doi.org/10.1016/S0140-6736(19)31721-0), indexed in Pubmed: 31378392.
21. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. 2019; 25(1): 70–74, doi: [10.1038/s41591-018-0240-2](https://doi.org/10.1038/s41591-018-0240-2), indexed in Pubmed: 30617318.
22. Attia ZI, Kapa S, Yao X, et al. Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. *J Cardiovasc Electrophysiol*. 2019; 30(5): 668–674, doi: [10.1111/jce.13889](https://doi.org/10.1111/jce.13889), indexed in Pubmed: 30821035.
23. Yao X, McCoy RG, Friedman PA, et al. ECG AI-Guided Screening for Low Ejection Fraction (EAGLE): Rationale and design of a pragmatic cluster randomized trial. *Am Heart J*. 2020; 219: 31–36, doi: [10.1016/j.ahj.2019.10.007](https://doi.org/10.1016/j.ahj.2019.10.007), indexed in Pubmed: 31710842.
24. Shah SJ, Katz DH, Selvaraj S, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*. 2015; 131(3): 269–279, doi: [10.1161/CIRCULATIONAHA.114.010637](https://doi.org/10.1161/CIRCULATIONAHA.114.010637), indexed in Pubmed: 25398313.
25. Katz DH, Deo RC, Aguilar FG, et al. Phenomapping for the Identification of Hypertensive Patients with the Myocardial Substrate for Heart Failure with Preserved Ejection Fraction. *J Cardiovasc Transl Res*. 2017; 10(3): 275–284, doi: [10.1007/s12265-017-9739-z](https://doi.org/10.1007/s12265-017-9739-z), indexed in Pubmed: 28258421.

26. Krittanawong C, Zhang H, Wang Z, et al. Artificial Intelligence in Precision Cardiovascular Medicine. *J Am Coll Cardiol*. 2017; 69(21): 2657–2664, doi: [10.1016/j.jacc.2017.03.571](https://doi.org/10.1016/j.jacc.2017.03.571), indexed in Pubmed: 28545640.
27. Wang L, Tong L, Davis D, et al. The application of unsupervised deep learning in predictive models using electronic health records. *BMC Med Res Methodol*. 2020; 20(1): 37, doi: [10.1186/s12874-020-00923-1](https://doi.org/10.1186/s12874-020-00923-1), indexed in Pubmed: 32101147.
28. Madani A, Arnaout R, Mofrad M, et al. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit Med*. 2018; 1, doi: [10.1038/s41746-017-0013-1](https://doi.org/10.1038/s41746-017-0013-1), indexed in Pubmed: 30828647.
29. Sengupta PP, Huang YM, Bansal M, et al. Cognitive Machine-Learning Algorithm for Cardiac Imaging: A Pilot Study for Differentiating Constrictive Pericarditis From Restrictive Cardiomyopathy. *Circ Cardiovasc Imaging*. 2016; 9(6), doi: [10.1161/CIRCIMAGING.115.004330](https://doi.org/10.1161/CIRCIMAGING.115.004330), indexed in Pubmed: 27266599.
30. Zhang J, Gajjala S, Agrawal P, et al. Fully Automated Echocardiogram Interpretation in Clinical Practice. *Circulation*. 2018; 138(16): 1623–1635, doi: [10.1161/CIRCULATIONAHA.118.034338](https://doi.org/10.1161/CIRCULATIONAHA.118.034338), indexed in Pubmed: 30354459.
31. Wolterink JM, Leiner T, de Vos BD, et al. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Med Image Anal*. 2016; 34: 123–136, doi: [10.1016/j.media.2016.04.004](https://doi.org/10.1016/j.media.2016.04.004), indexed in Pubmed: 27138584.
32. Coenen A, Kim YH, Kruk M, et al. 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, doi: [10.1161/CIRCIMAGING.117.007217](https://doi.org/10.1161/CIRCIMAGING.117.007217), indexed in Pubmed: 29914866.
33. Tesche C, De Cecco CN, Baumann S, et al. Coronary CT angiography-derived fractional flow reserve: machine learning algorithm versus computational fluid dynamics modeling. *Radiology*. 2018; 288(1): 64–72, doi: [10.1148/radiol.2018171291](https://doi.org/10.1148/radiol.2018171291), indexed in Pubmed: 29634438.
34. Motwani M, Dey D, Berman DS, et al. 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*. 2017; 38(7): 500–507, doi: [10.1093/eurheartj/ehw188](https://doi.org/10.1093/eurheartj/ehw188), indexed in Pubmed: 27252451.
35. Wolterink JM, Leiner T, Viergever MA, et al. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Trans Med Imaging*. 2017; 36(12): 2536–2545, doi: [10.1109/TMI.2017.2708987](https://doi.org/10.1109/TMI.2017.2708987), indexed in Pubmed: 28574346.
36. Zreik M, Lessmann N, van Hamersvelt RW, et al. Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis. *Med Image Anal*. 2018; 44: 72–85, doi: [10.1016/j.media.2017.11.008](https://doi.org/10.1016/j.media.2017.11.008), indexed in Pubmed: 29197253.
37. Haro Alonso D, Wernick MN, Yang Y, et al. Prediction of cardiac death after adenosine myocardial perfusion SPECT based on machine learning. *J Nucl Cardiol*. 2019; 26(5): 1746–1754, doi: [10.1007/s12350-018-1250-7](https://doi.org/10.1007/s12350-018-1250-7), indexed in Pubmed: 29542015.
38. Betancur J, Commandeur F, Motlagh M, et al. Deep learning for prediction of obstructive disease from fast myocardial perfusion SPECT: a multicenter study. *JACC Cardiovasc Imaging*. 2018; 11(11): 1654–1663, doi: [10.1016/j.jcmg.2018.01.020](https://doi.org/10.1016/j.jcmg.2018.01.020), indexed in Pubmed: 29550305.
39. Betancur J, Hu LH, Commandeur F, et al. Deep learning analysis of upright-supine high-efficiency SPECT myocardial perfusion imaging for prediction of obstructive coronary artery disease: a multicenter study. *J Nucl Med*. 2019; 60(5): 664–670, doi: [10.2967/jnumed.118.213538](https://doi.org/10.2967/jnumed.118.213538), indexed in Pubmed: 30262516.
40. Betancur J, Otaki Y, Motwani M, et al. Prognostic value of combined clinical and Myocardial perfusion imaging data using machine learning. *JACC Cardiovasc Imaging*. 2018; 11(7): 1000–1009, doi: [10.1016/j.jcmg.2017.07.024](https://doi.org/10.1016/j.jcmg.2017.07.024), indexed in Pubmed: 29055639.
41. Bai W, Sinclair M, Tarroni G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson*. 2018; 20(1): 65, doi: [10.1186/s12968-018-0471-x](https://doi.org/10.1186/s12968-018-0471-x), indexed in Pubmed: 30217194.
42. Fries JA, Varma P, Chen VS, et al. Weakly supervised classification of aortic valve malformations using unlabeled cardiac MRI sequences. *Nat Commun*. 2019; 10(1): 3111, doi: [10.1038/s41467-019-11012-3](https://doi.org/10.1038/s41467-019-11012-3), indexed in Pubmed: 31308376.
43. Yang G, Zhuang X, Khan H, et al. Fully automatic segmentation and objective assessment of atrial scars for long-standing persistent atrial fibrillation patients using late gadolinium-enhanced MRI. *Med Phys*. 2018; 45(4): 1562–1576, doi: [10.1002/mp.12832](https://doi.org/10.1002/mp.12832), indexed in Pubmed: 29480931.
44. Swift AJ, Lu H, Uthoff J, et al. A machine learning cardiac magnetic resonance approach to extract disease features and automate pulmonary arterial hypertension diagnosis. *Eur Heart J Cardiovasc Imaging*. 2020 [Epub ahead of print], doi: [10.1093/ehjci/jeaa001](https://doi.org/10.1093/ehjci/jeaa001), indexed in Pubmed: 31998956.
45. Yang Su, Kweon J, Roh JH, et al. Deep learning segmentation of major vessels in X-ray coronary angiography. *Sci Rep*. 2019; 9(1): 16897, doi: [10.1038/s41598-019-53254-7](https://doi.org/10.1038/s41598-019-53254-7), indexed in Pubmed: 31729445.
46. Kwon JM, Kim KH, Akkus Z, et al. Artificial intelligence for detecting mitral regurgitation using electrocardiography. *J Electrocardiol*. 2020; 59: 151–157, doi: [10.1016/j.jelectrocard.2020.02.008](https://doi.org/10.1016/j.jelectrocard.2020.02.008), indexed in Pubmed: 32146201.
47. Weng SF, Reys J, Kai J, et al. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*. 2017; 12(4): e0174944, doi: [10.1371/journal.pone.0174944](https://doi.org/10.1371/journal.pone.0174944), indexed in Pubmed: 28376093.
48. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med*. 2018; 1: 18, doi: [10.1038/s41746-018-0029-1](https://doi.org/10.1038/s41746-018-0029-1), indexed in Pubmed: 31304302.
49. Brajer N, Cozzi B, Gao M, et al. Prospective and External Evaluation of a Machine Learning Model to Predict In-Hospital Mortality of Adults at Time of Admission. *JAMA Netw Open*. 2020; 3(2): e1920733, doi: [10.1001/jamanetworkopen.2019.20733](https://doi.org/10.1001/jamanetworkopen.2019.20733), indexed in Pubmed: 32031645.
50. Shameer K, Johnson KW, Yahi A, et al. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using mount sinai heart failure cohort. *Pac Symp Biocomput*. 2017; 22: 276–287, doi: [10.1142/9789813207813_0027](https://doi.org/10.1142/9789813207813_0027), indexed in Pubmed: 27896982.
51. Shouval R, Hadanny A, Shlomo N, et al. Machine learning for prediction of 30-day mortality after ST elevation myocardial infarction: an acute coronary syndrome israeli survey data mining study. *Int J Cardiol*. 2017; 246: 7–13, doi: [10.1016/j.ijcard.2017.05.067](https://doi.org/10.1016/j.ijcard.2017.05.067), indexed in Pubmed: 28867023.
52. Wallert J, Tomasoni M, Madison G, et al. Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data. *BMC Med Inform Decis Mak*. 2017; 17(1): 99, doi: [10.1186/s12911-017-0500-y](https://doi.org/10.1186/s12911-017-0500-y), indexed in Pubmed: 28679442.

53. Pieszko K, Hiczkiewicz J, Budzianowski P, et al. Predicting long-term mortality after acute coronary syndrome using machine learning techniques and hematological markers. *Dis Markers*. 2019; 2019: 9056402, doi: [10.1155/2019/9056402](https://doi.org/10.1155/2019/9056402), indexed in Pubmed: [30838085](https://pubmed.ncbi.nlm.nih.gov/30838085/).
54. Pieszko K, Hiczkiewicz J, Budzianowski P, et al. Machine-learned models using hematological inflammation markers in the prediction of short-term acute coronary syndrome outcomes. *J Transl Med*. 2018; 16(1): 334, doi: [10.1186/s12967-018-1702-5](https://doi.org/10.1186/s12967-018-1702-5), indexed in Pubmed: [30509300](https://pubmed.ncbi.nlm.nih.gov/30509300/).
55. Gibson WJ, Nafee T, Travis R, et al. Machine learning versus traditional risk stratification methods in acute coronary syndrome: a pooled randomized clinical trial analysis. *J Thromb Thrombolysis*. 2020; 49(1): 1–9, doi: [10.1007/s11239-019-01940-8](https://doi.org/10.1007/s11239-019-01940-8), indexed in Pubmed: [31535314](https://pubmed.ncbi.nlm.nih.gov/31535314/).
56. Kalscheur MM, Kipp RT, Tattersall MC, et al. Machine learning algorithm predicts cardiac resynchronization therapy outcomes: lessons from the COMPANION trial. *Circ Arrhythm Electrophysiol*. 2018; 11(1): e005499, doi: [10.1161/CIRCEP.117.005499](https://doi.org/10.1161/CIRCEP.117.005499), indexed in Pubmed: [29326129](https://pubmed.ncbi.nlm.nih.gov/29326129/).
57. Panch T, Mattie H, Celi L. The “inconvenient truth” about AI in healthcare. *npj Digital Medicine*. 2019; 2(1), doi: [10.1038/s41746-019-0155-4](https://doi.org/10.1038/s41746-019-0155-4), indexed in Pubmed: [31453372](https://pubmed.ncbi.nlm.nih.gov/31453372/).
58. Oakden-Rayner L, Dunnmon J, Carneiro G, et al. Hidden stratification causes clinically meaningful failures in machine learning for medical imaging. *Proceedings of the ACM Conference on Health, Inference, and Learning*. 2020, doi: [10.1145/3368555.3384468](https://doi.org/10.1145/3368555.3384468).
59. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017; 542(7639): 115–118, doi: [10.1038/nature21056](https://doi.org/10.1038/nature21056), indexed in Pubmed: [28117445](https://pubmed.ncbi.nlm.nih.gov/28117445/).
60. Goodman B, Flaxman S. European Union Regulations on Algorithmic Decision-Making and a “Right to Explanation”. *AI Magazine*. 2017; 38(3): 50–57, doi: [10.1609/aimag.v38i3.2741](https://doi.org/10.1609/aimag.v38i3.2741).
61. Gale W, Oakden-Rayner L, Carneiro G, et al. Producing radiologist-quality reports for interpretable deep learning. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). 2019, doi: [10.1109/isbi.2019.8759236](https://doi.org/10.1109/isbi.2019.8759236).
62. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: an attentional integration. *Hum Factors*. 2010; 52(3): 381–410, doi: [10.1177/0018720810376055](https://doi.org/10.1177/0018720810376055), indexed in Pubmed: [21077562](https://pubmed.ncbi.nlm.nih.gov/21077562/).
63. Hripacsak G, Knirsch C, Zhou Li, et al. Bias associated with mining electronic health records. *J Biomed Discov Collab*. 2011; 6: 48–52, doi: [10.5210/disco.v6i0.3581](https://doi.org/10.5210/disco.v6i0.3581), indexed in Pubmed: [21647858](https://pubmed.ncbi.nlm.nih.gov/21647858/).
64. Gianfrancesco MA, Tamang S, Yazdany J, et al. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med*. 2018; 178(11): 1544–1547, doi: [10.1001/jamainternmed.2018.3763](https://doi.org/10.1001/jamainternmed.2018.3763), indexed in Pubmed: [30128552](https://pubmed.ncbi.nlm.nih.gov/30128552/).
65. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. *JAMA*. 2019 [Epub ahead of print], doi: [10.1001/jama.2019.18058](https://doi.org/10.1001/jama.2019.18058), indexed in Pubmed: [31755905](https://pubmed.ncbi.nlm.nih.gov/31755905/).
66. Gur D, Bandos A, Cohen C, et al. The “laboratory” effect: comparing radiologists’ performance and variability during prospective clinical and laboratory mammography interpretations. *Radiology*. 2008; 249(1): 47–53, doi: [10.1148/radiol.2491072025](https://doi.org/10.1148/radiol.2491072025), indexed in Pubmed: [18682584](https://pubmed.ncbi.nlm.nih.gov/18682584/).
67. Lehman CD, Wellman RD, Buist DSM, et al. Breast Cancer Surveillance Consortium. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med*. 2015; 175(11): 1828–1837, doi: [10.1001/jamainternmed.2015.5231](https://doi.org/10.1001/jamainternmed.2015.5231), indexed in Pubmed: [26414882](https://pubmed.ncbi.nlm.nih.gov/26414882/).
68. Angus DC. Randomized clinical trials of artificial intelligence. *JAMA*. 2020 [Epub ahead of print], doi: [10.1001/jama.2020.1039](https://doi.org/10.1001/jama.2020.1039), indexed in Pubmed: [32065828](https://pubmed.ncbi.nlm.nih.gov/32065828/).
69. Wang Pu, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*. 2019; 68(10): 1813–1819, doi: [10.1136/gutjnl-2018-317500](https://doi.org/10.1136/gutjnl-2018-317500), indexed in Pubmed: [30814121](https://pubmed.ncbi.nlm.nih.gov/30814121/).
70. Wijnberge M, Geerts BF, Hol L, et al. Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the HYPE randomized clinical trial. *JAMA*. 2020 [Epub ahead of print], doi: [10.1001/jama.2020.0592](https://doi.org/10.1001/jama.2020.0592), indexed in Pubmed: [32065827](https://pubmed.ncbi.nlm.nih.gov/32065827/).
71. Schwab K. *The fourth industrial revolution*. Currency. 2017: 192.