Artificial Intelligence in Patients with Congenital Heart Disease: Where Do We Stand?

EDITOR'S PICK

The authors have performed a literature search of how state-of-the-art imaging diagnoses for congenital heart disease are taking place with the help of artificial intelligence (AI). There is a clinical unmet need for the use of AI in congenital heart disease, and a joint effort is needed to spread AI knowledge and applications in the medical field to improve diagnosis, treatment, and outcomes for patients. This review gives an overview of AI usage in diagnostic imaging, electrocardiograms, and clinical diagnosis in patients with congenital heart disease.

Authors: *Marinka D. Oudkerk Pool,[1,2] Dirkjan Kauw,[1,2] Hidde Bleijendaal,[1] Barbara J.M. Mulder,[1] Yigal Pinto,[1] Berto J. Bouma,[1] Michiel M. Winter[1]

1. Department of Cardiology, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, the Netherlands
2. Netherlands Heart Institute, Utrecht, the Netherlands

*Correspondence to m.d.oudkerkpool@amsterdamumc.nl

Disclosure: The authors have declared no conflicts of interest.

Received: 17.02.2020

Accepted: 30.03.2020

Keywords: artificial intelligence (AI), congenital heart disease (CHD), machine learning (ML), deep learning (DL).

Citation: EMJ Cardiol. 2020;8[1]:70-81.

Abstract

Life expectancy of patients with congenital heart disease (CHD) has increased in recent decades; however, late complications remain frequent and difficult to predict. Progress in data science has spurred the development of decision support systems and could aid physicians in predicting clinical deterioration and in the management of CHD patients. Newly developed artificial intelligence (AI) algorithms have shown performances comparable to humans in clinical diagnostics using statistical and computational algorithms and are expected to partly surpass human intelligence in the near future. Although much research on AI has been performed in patients with acquired heart disease, little data is available with respect to research on AI in patients with CHD. Learning algorithms in patients with CHD have shown to be promising in the interpretation of ECG, cardiac imaging, and the prediction of surgical outcome. However, current learning algorithms are not accurate enough to be implemented into daily clinical practice. Data on AI possibilities remain scarce in patients with CHD, and studies on large data sets are warranted to increase sensitivity, specificity, accuracy, and clinical relevance of these algorithms.

INTRODUCTION

Improved medical treatment and surgical techniques has caused the life expectancy of patients with congenital heart disease (CHD) to be significantly prolonged.[1-3] As these patients reach adulthood, late complications such as arrhythmias and congestive heart failure occur,[1] resulting in reduced quality of life and life expectancy.[4] Furthermore, these complications often result in unscheduled hospital visits or even emergency admissions.[3,5-7] Although visits to the outpatient clinic are frequent, it remains difficult to predict and prevent clinical deterioration.

With the introduction of the electronic medical record and the ability to digitally store data for diagnostic modalities, such as ECG and echocardiography, large amounts of patient data have been generated over the past few decades. Using machine learning (ML) or deep learning (DL) for the analysis of these data has been a topic of interest for some years. Progress in data science has spurred the development of decision support systems which can aid physicians in the management of CHD patients and therapeutic decision making.[4,8,9] Newly developed algorithms perform as well as humans in clinical diagnostics using statistical and computational algorithms to perform recognition, classification, and learning tasks, and are expected to outperform humans in the near future.[10,11]

Used in this context, artificial intelligence (AI) is an umbrella term for the use of computers to model intelligent behaviour. The terms 'neural networks,' DL, and ML are technical concepts that fall under this umbrella but are confusingly used interchangeably with the term AI in the literature. This review will focus on ML and DL. Learning algorithms learn from data given as an input, also called the training dataset. The algorithm then gets tested on a so-called test or validation dataset, which contains new unseen data. ML, specifically, refers to algorithms that can 'learn' patterns from training data and then use the learned patterns to classify previously unseen data.[12] DL algorithms have an extra hidden layer, which allows them to automatically detect important features from the data, while in ML algorithms, the features need to be provided manually.[4,13]

Although several AI studies have been performed in patients with CHD over the past few years, data are limited compared to data in patients with acquired heart disease.[4,8,9,13-16] Considering the propensity of CHD patients for developing arrhythmias and heart failure, the predictive abilities of the AI algorithms could prove to be lifesaving. Therefore, the aim of this review is to provide an overview of studies investigating the potential of AI algorithms with respect to the various imaging modalities in patients with CHD.

METHODS

Literature Search

Medline® (Northfield, Illinois, USA) and EMBASE (Elsevier, Amsterdam, the Netherlands) were used to search for studies published up to 9th August 2019. The search was developed iteratively for synonyms of 'congenital heart disease,' ML, DL, and AI, both controlled vocabulary (Medical Subject Headings [MeSH]) and free-text words. Nonhuman studies, case reports, biomarker studies, and reviews were excluded. The reference list and cited articles were checked for additional references.

Selection of Studies

Studies were included if they applied AI algorithms for diagnostics (heart sound, echocardiography, MRI, CT, electrocardiogram analysis, and classification/prediction models, for example) in CHD patients. Since the terms AI, DL, and ML are used interchangeably, all three terms were included in this review. All potential articles were read in full by two authors (Ms Marinka D. Oudkerk Pool and Mr Dirkjan Kauw). Disagreements concerning eligibility were resolved by discussion.

Extraction of Data

The extracted data from each paper were author, publication year, total number of patients (both training and test set), patient population, data used for analysis (input data in the algorithm), primary outcome (goal of the study), the used AI algorithm, and accuracy of the proposed AI algorithm. For comparison between the different techniques the sensitivity (SE), specificity (SP), and accuracy were used.

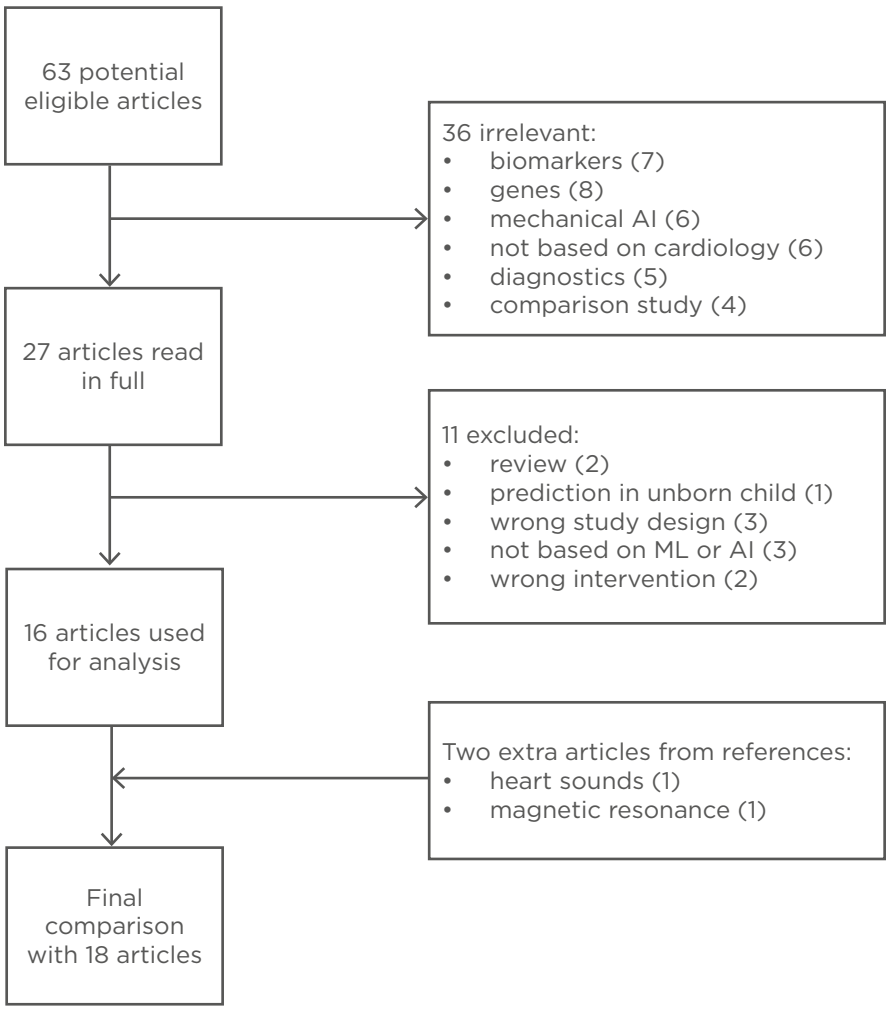

Figure 1: Flow diagram of search query.

AI: artificial intelligence; ML: machine learning.

Accuracy is defined as the number of correctly classified results compared to the 'true' value (either positive or negative), as assessed by the gold standard technique.

True positive (TP) is the proportion of actual positives that are correctly identified as such. True negative (TN) is the proportion of actual negatives correctly identified as such. False positive (FP) is a negative value identified as a positive value, and false negative (FN) is a positive value identified as a negative value. SE, SP, and accuracy can be defined using equations:[17-20]

$$SE = 100 \times \frac{TP}{TP + FN}$$

$$SE = 100 \times \frac{TP}{TN + FP}$$

$$\text{accuracy} = 100 \times \frac{TP + TN}{TP+TN+FP+FN}$$

RESULTS

In total, 63 articles were potentially eligible for this review after removing duplicates. Forty-eight articles were considered irrelevant because they focussed on biomarkers, genes, mechanical AI, were not based on cardiology (either neurology or mechanical ventilation), diagnostics (over the phone or medication), or comparison study in which two or more imaging modalities were compared. Twenty-seven articles were read in full, after which an additional 11 articles were excluded. Two additional articles were selected by going through the references. The final analysis consisted of 18 articles (Figure 1).

Topics of these studies included analyses of cardiac imaging modalities (echocardiography, MRI), ECG, and clinical prediction models using AI algorithms.

Table 1: Selected articles.

First author (year)	n (training set)	n (test set)	Patient population	Category for analysis	Data used for analysis	Primary outcome	Learning algorithm	Accuracy (%)
Elgendi et al.[21] (2015)	27. Unclear how data has been split.		Patients who were undergoing right heart catheterisation	Heart sound	Heart sounds measured with 3M™ Littmann® 3200 Electronic Stethoscope*	Recorded heart sounds to distinguish subjects with PAH	Linear discriminant analysis	SE: 92.86 SP: 92.31
Gharehbaghi et al.[22] (2015)	50. Unclear how data has been split.		Twenty-eight healthy children and 22 children with BAV	Heart sound	Heart sounds measured with Meditron Electronic Stethoscope†	Develop algorithm for detecting BAV in children	Support vector machine	72.9 Practitioner: 71.60
Gharehbaghi et al.[9] (2017)	90. Unclear how data has been split.		Fifty-five healthy children and 35 children with BAV	Heart sound	Heart sounds measured with Meditron Electronic Stethoscope	Detecting BAV, healthy, or MR from heart sound	Combination of hidden Markolov model and support vector machine	86.4
Elgendi et al.[23] (2018)	60. Unclear how data has been split.		Patients who were undergoing right heart catheterisation	Heart sound	Heart sounds measured with 33M Littmann 3200 Electronic Stethoscope	Recorded heart sounds to distinguish subjects with PAH	Linear discriminant analysis	SE: 84.00 SP: 88.57
DeGroff et al.[24] (2001)	69 used for both training and validation		Paediatric patients	Heart sound	Heart sounds measured with Cambridge Heart Sound Microphone	Distinguish between innocent and pathological	Artificial neural network	SE: 100.00 SP: 100.00
Sepehri et al.[25] (2009)	60	60	Heart sounds from database (both healthy and with CHD)	Heart sound	Phonocardiogram and electrocardiogram	Identifying children with congenital heart disease	Artificial neural network	93.6
Thompson et al.[26] (2018)	603. Unclear how data has been split.		Heart sounds from database (John Hopkins Outpatient Center, Baltimore, Maryland)	Heart sound	Heart sounds recorded with an electronic stethoscope with corresponding ECG	Distinguish between innocent and pathological	Not specified	88
Bhatikar et al.[27] (2004)	241. Unclear how data has been split.		Heart sounds from database (The Children's Hospital, Denver, Colorado)	Heart sound	Heart sounds from a microphone optimised for low frequencies	Diagnosis of heart murmurs in paediatrics	Artificial neural network	SE: 83.00 SP: 90.00

Table 1 continued

First author (year)	n (training set)	n (test set)	Patient population	Category for analysis	Data used for analysis	Primary outcome	Learning algorithm	Accuracy (%)
Sepehri et al.[28] (2016)	134	129	Children referred to the hospital	Heart sound	Phonocardiogram and electrocardiogram	Identifying children with congenital heart disease	Arash-Band	87.45
Diller et al.[4] (2018)	159	40	Patients undergoing routine transthoracic examinations	Ultrasound	Echocardiographic data	Discriminate between patients with TGA after atrial switch operation, patients with ccTGA and normal controls	Convolution neural network	98
Pereira et al.[16] (2017)	163	91	Boston Children's Hospital database	Ultrasound	Echocardiographic data	Detecting CoA in newborns via ultrasound	Support vector machine	Not mentioned
Neukamm et al.[29] (2013)	30. Unclear about split between sets.		Patients with Tetralogy of Fallot after pulmonary valve replacement	Ultrasound	End-diastolic volume, end-systolic volume, and ejection fraction	Volumetric assessment compared to gold standard MRI	Knowledge-based	EDV; SE: 100.00, SP: 86.00 ESV; SE: 78.00, SP: 86.00 EF; SE: 75.00, SP: 43.00
Nyns et al.[30] (2016)	17. Unclear about split between sets.		Children and adolescents	MRI	Ventricular volumetry, using a short-axis cine stack	Evaluate feasibility, accuracy and labour intensity compared to conventional Simpson's method	Knowledge-based	EDV: 82.00 ESV: 93.00 EF: 73.00
Yang et al.[13] (2002)	106 used for both training and testing.		ECG from database (Nagoya University Hospital, Nagoya, Japan)	ECG	ECG features: waveforms, and voltages of upright and negative deflections	Differentiate between ASD and non-ASD	Artificial neural network	ACC: 91.50 SE: 91.40 SP: 91.70
Ruiz-Fernandez et al.[8] (2015)	2432 used for both training and testing.		Children heart disease database	Classification model	Presurgical and postsurgical data	Classifying the risk of paediatric cardiac surgery	Multilayer Perceptron; Radial Basis Function; Self-organising Map; Decision Tree	99.87; 95.60; 81.79; 80.09

Table 1 continued

First author (year)	n (training set)	n (test set)	Patient population	Category for analysis	Data used for analysis	Primary outcome	Learning algorithm	Accuracy (%)
Ruiz et al.[14] (2019)	93 patients. Unclear about split between sets.		Infants with single-ventricle physiology	Prediction model	Inpatient data	Predict critical events early and accurately	Naïve Bayes	SE: 84.00 SP: 81.00
Diller et al.[15] (2018)	8.015	2.004	Adult patients under active follow-up in Royal Brompton Hospital, London, UK	Classification model	Patient data, including diagnosis, clinical status, and medication	Categorise patients in diagnostic and disease complexity subgroups	Deep Learning	90.20
Chiogna et al.[31] (1996)	457	114	Neonates	Classification model	21 questions based on clinical presentation, blood gasses, and imaging modalities	27 congenital heart disease classes	Decision tree	59.00

*Saint Paul, Minnesota, USA.

†Welch Allyn®, Skaneateles Falls, New York, USA.

ACC: accuracy; ASD: atrial septal defect; BAV: bicuspid aortic valve; ccTGA: congenitally-corrected transposition of great arteries; CHD: congenital heart disease; CoA: coarctatio aortae; EDV: end-diastolic volume; EF: ejection fraction; ESV: end-systolic volume; MR: mitral regurgitation; PAH: pulmonary artery hypertension; SE: sensitivity; SP: specificity; TGA: transposition of great arteries.

Table 1 shows an overview of each selected article found in this search. No articles on CT in CHD patients were found during this search.[4,8,9,13-16,21-31]

Patient Population

In total, 15,244 patients were analysed: 10,354 adults (35% male; 33% female; 32% no gender described; mean age of 33.30 ±13.00 years), 1,858 children (>2 years old; mean age of 9.22 ±1.09 years [42% of patients]; 58% no age described), and 4,099 were infants (<2 years old; age not described). Diagnoses of CHD of varying complexity were made in 14,532 (95.33%) patients and 712 (4.67%) patients were included as a healthy control group.

Learning Algorithms

In the 18 analysed articles, 15 different AI algorithms were used. The most used technique was the artificial neural network (ANN) in four of the articles (22%). Table 2 gives an overview of all techniques used in this review. Nine out of 18 articles analysed the use of ML in heart sounds (50%). The other articles analysed echocardiography (n=3, 17%), MRI (n=1, 6%), ECG (n=1, 6%), as well as prediction or classification models (n=4, 22%).

Heart Sound Analysis

Nine articles aimed to distinguish between pathological or innocent murmurs using ML on sound recordings from an electronic stethoscope.

The study by DeGroff et al.[24] aimed to determine pathological from innocent murmurs using spectral resolution and frequency range as input. Using an ANN, they found high SP and SE (both >90.0%). Sepehri et al.[25] also found high accuracy (93.6%) using an ANN based on spectral and timing properties of the sound recordings of heart sounds of murmurs. The algorithm was trained on 60 normal and 60 pathological heart sound recordings.

Table 2: Overview of artificial intelligence algorithms used in the review.

Support vector machine (SVM)	The SVM searches for the best separation line or decision boundary (also called the maximum margin hyperplane) between two groups of data. The maximum margin is the equidistance between the closest vectors of both groups. These two vectors are called the support vectors. The maximum margin hyperplane is determined by the training dataset. Any new vector from the testing or validation dataset will fall on one side of the hyperplane or the other; based on this, the vector is sorted into the correct class.[9,25]
Decision tree (DT)	In a DT, the algorithm is dividing the data into different groups. During each separation, the algorithm tries to maximise the number of vectors from a certain category. Rules for separation are generated by following the path from branch to leaf; however, since a lot of paths are possible, it can reach a considerable size. DT are not very powerful on their own but can be used in other methods that leverage their simplicity.[8]
Knowledge-based	The knowledge-based algorithm tries to reconstruct shapes of the heart based on anatomical landmarks. The database already contains information of reconstructed 3D surface models of the preferred anatomy. The algorithm will reconstruct the requested anatomy from a new patient based on data out of the database.[30]
Arash-band method	Every heart disease has a discriminative frequency band, named the Arash-Band. The Arash-Band is defined as the spectral energy band that provides maximum discrimination with respect to the normal condition. The Arash-Band is calculated during the training phase, using statistical techniques as the discriminating tools for the band selection.[28]
Linear discriminant analysis (LDA)	The LDA tries to find a linear combination of features to separate between two or more classes. The LDA has continuous independent variables (the features) and the class label as a continuous dependent variable.[21]
Hidden Markov model (HMM)	A Markov model can be used to calculate a probability for an observable event, an HMM also looks at hidden events, such as part-of-speech tags. In an HMM there is a hidden Markov layer, which contains a Markov chain. A Markov chain is a model that calculates the probabilities of sequences of random variables. A Markov chain only takes into account the current state and does not account for any previous state.[9]
Self-organising map (SOM)	This algorithm represents competitive learning, all neurons compete to be the closest to the input value. The Euclidian distance is used to measure similarities between the input value and each neurons' weight in order to choose the winning neuron. Afterwards, the weight of the winner and its neighbours is updated for the next input value.[8]
Naïve Bayes (NB)	The NB assumes that every feature is independent of the value of any other feature. The classifier is based on Bayes theorem, meaning the probability of A happening, given B has occurred is equal to the probability of B happening, given A has occurred times the probability of A, divided by the probability of B. The NB is an oversimplification of assumptions but tends to work quite well in complex problems.[14]
Deep learning (DL)	Deep Learning refers to every algorithm with multiple layers, making it deep. In the article by Diller et al.[15] it is referred to a combination of convolutional network and a dense network. A dense network is a network in which the number of links between nodes is close to the maximal number of nodes.[15]
Artificial neural network (ANN)	The neural networks are made to mimic the neural networks in the brain. The neural network usually consists of an input layer, one or more hidden layers, and an output layer. The input values represent the dendrites going to a neuron, which then ends in the output signal, representing the axon. The neuron determines if the sum of the input values is important and if they get passed along to the output value. The neuron is represented by the so-called activation function. The weighted sum of the input values is applied in the activation function (most common functions are the threshold, sigmoid, rectifier, or hyperbolic tangent function). Every activation function has a different method to determine how it will be activated. If the weighted sum is high enough, the signal will be passed onto the output value. Every input value is connected to every hidden layer neuron, from which one value is passed to the output layer neuron, after which one value is the output value. The output value can be binary, categorical, or continuous. The predicted output value is compared to the actual output value. Using a cost-function, the error between the output values is compared, this error is fed back to the neuron, and the weights of the input are updated (this is called backpropagation). This process is repeated for all the values of the training dataset. After which the final weights are applied to the test or validation dataset.[24,27]

Table 2 continued

Convolutional neural network (CNN)	CNN have a similar algorithm to the ANN, however the input for a CNN is an image as to a value in the ANN. The image gets down sampled by a feature detector (also known by kernel or filter), which results in a feature map. The size of the feature detector determines the size of the feature map. The bigger the feature detector, the smaller the feature map will be and the bigger the down-sampling of the original image. Many feature maps will be created to obtain the first convolutional layer, using different feature detectors. Trough training the algorithm determines how the feature detectors should look to preserve important features from the original image. After the convolutional layer, a rectified linear unit function is applied to remove all linearity since images themselves are highly nonlinear. When making the feature maps, linearity can occur due to the down sampling. Then a pooling layer has been applied. The pooling layer makes the image recognisable even if the image has been tilted or shifted and makes overfitting less possible by reducing the size of the convolutional layer. Lastly, a flattening layer conducts a similar function as the activation function in the ANN. After the flattening layer, the output of this process will be used as an input for an algorithm that is similar to the ANN. In the backpropagation not only are the weights adjusted, but also the feature detectors, creating a more accurate down-sampling of the image.[4]
Radial basis function networks (RBF)	RBF is a type of ANN. The input layer and the hidden layer do not have associated weights. Each neuron in the hidden layer represents an RBF. The neurons compute the Euclidean distance between the synaptic weights vector and the input values. Over this distance the RBF is applied, which most often is the Gaussian function.[8]
Multi-layer perceptron (MLP)	The MLP is an example of feedforward ANN. In general, it refers to multiple layers of perceptrons with threshold activation.[8,25]

To evaluate the algorithm, it was tested with 60 either innocent or pathological murmurs to correctly identify first and second heart sounds. Other articles used multiple algorithms to distinguish between a pathological or innocent murmur, namely linear discriminant analysis, support vector machine, a combination between hidden Markov model and support vector machine, ANN, and the Arash-band. ANN was the most frequently used algorithm (n=3, 33%), and yielded the highest accuracy, SE, and SP.

Echocardiographic Analysis

AI algorithms were used on echocardiographic data to distinguish between structurally normal or pathological hearts, or to determine cardiac cavity volumes and function. The algorithm can be trained to detect change in echogenicity in the collected data, which can be seen in the wall of the heart. In this manner, Diller et al.[4] found accuracy of 98% in distinguishing between transposition of the great arteries (TGA) after an atrial switch operation, congenitally corrected-TGA, and normal controls using a convolutional neural network (CNN) algorithm. The endocardial border was marked by two researchers and compared to the border marked by a CNN algorithm. A knowledge-based article written by Neukamm et al.[29] only looked at SE and SP and found that making a 3D model out of the 2D echocardiogram data is feasible in 97% of cases; however, results for assessing the ejection fraction (EF) were unsatisfactory and MRI remains the method of choice.

MRI

In one of the articles by Nyns et al.,[30] MRI was used as an input for knowledge-based reconstruction of the volume of the right ventricle after atrial switch operation in patients with a TGA. In a knowledge-based reconstruction, the input is compared to a database that contains information on the 3D model of the place of interest and tries to reconstruct based on this database.[30] The knowledge-based reconstruction was compared to the gold standard, which is the Simpson's method. The Simpson's method is a geometric model in which the right ventricle is calculated based on the sum of a cylinder (base of the heart to the tricuspid valve).[32] The accuracy of the end-diastolic volume (82%), end-systolic volume (93%), and EF (73%) were compared. Knowledge-based reconstruction is a feasible, accurate, and fast method compared to the gold standard for measuring right ventricle volumes in patients after arterial switch operation.[30]

Electrocardiogram Analysis

One article was found using ML on ECG of patients with CHD. In this article by Yang et al.,[13] the authors aimed to distinguish atrial septum defect from patients with nonatrial septum defect and healthy controls' ECG. The QRS and T wave measurements from lead I, lead II, and all precordial leads were used as input. A SE of 91.4% and SP of 91.7% was found, with an accuracy of 91.5% using an ANN.

Classification Model

In the classification and prediction models, the ML algorithms were used to predict clinical deterioration, to classify surgical risk, or to classify the heart disease using patient characteristics. If the output of the network is categorical, it will make a prediction model. If the output has discrete values, the algorithm will do a classification of the data.[33] Ruiz-Fernandez et al.[8] found an accuracy of 99.9% in classifying the risk of mortality in paediatric surgery using the multilayer perceptron algorithm. The goal of this study was to develop a clinical decision support system to help cardiologists decide whether surgery was indicated. Ruiz et al.[14] investigated early prediction of critical events in infants using a naïve Bayesian model. Thirty-four routinely collected data points, such as heart rate, CO_2, and lactate, were used as input for the models. The model was able to detect future events up to 1 hour away with a SE of 84.0% and a SP of 81.0%. Diller et al.[15] used DL techniques (statistical learning which extracts features from raw data) to categorise diagnostic group, disease complexity, and New York Heart Association (NYHA) class, with an accuracy of 90.2%. In addition, they also estimated prognosis of the disease of adults with all types of CHD and to decide if patients needed to be discussed in the multidisciplinary team. Lastly, Chiogna et al.[31] used a decision tree algorithm to classify neonates with CHD into 27 disease classes, compared to an expert opinion. Input data consisted of routinely clinical data acquired at birth, such as ECG data, pO_2, heart size based on the chest X-ray, partial pressure of CO_2, and oligemic lung fields. Accuracy of 59.0% was achieved.

Discussion

This review provides an overview of the possibilities of AI for patients with CHD. Although AI algorithms have been used for patients since 2001, relatively few articles have been published on this subject. However, AI algorithms are gaining popularity in healthcare and especially in cardiology.[24] This is also demonstrated in this review since most articles are of relatively recent date (earliest dated 2015). In this review the authors found high SE and SP in most categories (echocardiographic data, ECG data, and in prediction/classification models), which means AI algorithms have great potential as an additional diagnostic tool in patients with CHD. However, the SE, SP, and accuracy are not yet high enough to be able to implement these algorithms safely in daily practice.

Most of the articles used ML on heart sounds, with high SP and SE. The highest accuracy (94%) was found using the ANN algorithm. Heart sound analysis is noninvasive, inexpensive to perform, and remains an important diagnostic tool in both adults and children. Overall, the techniques that were used distinguished between healthy and pathological sounds only, which might be useful as a primary screening tool. Heart sound analysis in patients with acquired heart disease also showed high SE, SP, and accuracy.[34] Ari et al.[35] managed to distinguish between aortic insufficiency, aortic stenosis, atrial septal defect, mitral regurgitation, mitral stenosis, or normal heart sound with an accuracy of 92%. These techniques can establish a diagnosis but do not yet determine the severity of the valve lesion. However, one could argue that heart sound analysis using ML should not be the main objective, as other noninvasive methods (echocardiography and cardiovascular MRI) are likely to be more informative if interpreted by ML techniques. The technique could be used as a screening tool by general practitioners to distinguish who should be sent to the hospital for further check-up.

Learning algorithms on noninvasive cardiac imaging (echocardiographic and MRI) shows a high accuracy when using a CNN algorithm, especially in the assessment of cardiac volumes.[4] AI algorithms have been used for echocardiographic imaging since 2006, but only started gaining popularity since 2012. Asch

et al.[36] trained a ML algorithm to automatically estimate the left ventricular EF on a database of >50 echocardiographic studies, including the apical 2- and 4-chamber views, and were compared to the left ventricular EF as assessed by the echocardiographist or cardiologist. The ML algorithm proved less sensitive (90% versus 93%), but more specific (92% versus 87%), and accurate (92% versus 89%), which makes the algorithms highly feasible in daily practice. However, patients with CHD tend to develop problems in their right ventricle. Genovese et al.[37] analysed 3D quantification of the right ventricle size and function in 56 patients receiving both cardiac magnetic resonance and 3D echocardiography exam on the same day. Echocardiographic volumes were analysed with a ML technique and compared with the cardiac MRI using the Bland-Altman and linear regression analyses. The automated ML analysis was correct in 18 patients (32%) but needed corrections in the remaining 38 patients. Although an intraclass correlation coefficient of 97% could be reached for the end-diastolic volume, 98% for the end-systolic volume, and 95% for the EF, the accuracy of a ML algorithm remains strongly correlated with the image quality. It seems likely that with increasing image quality, ML algorithms for its interpretation will become more reliable.

As with the echocardiography, cardiac MRI combined with ML has been gaining popularity and is already being used in daily practice. Ruijsink et al.[38] tried to analyse the cardiac magnetic resonance imaging using DL algorithms to automate ventricular function assessment for both ventricles, and reached SE of 95%, SP of 83%, and an accuracy of 89%. In the technique described by Nyns et al.,[30] the right ventricle was automatically analysed, but the key anatomical landmarks needed to be selected beforehand. More research is needed to evaluate if the DL algorithm could also make this process quicker with comparable SE, SP, and accuracy.

Remarkably, the sole article on ECG in CHD patients dates from 2002, although a lot of ML is conducted on ECG data in the general cardiac population[33,39-43]. Adult patients with CHD of 10 experience arrhythmias, which makes them a suitable group to use ML techniques with to predict events.[44] However, the baseline ECG recordings of these patients often already has an abnormal appearance and differs between patients with the same congenital heart defect, further complicating the analysis of the ECG. In patients with acquired heart disease, the most analysed arrhythmia is atrial fibrillation. Using ML, ECG characteristics during sinus rhythm can be determined to establish the presence of an atrial fibrillation signature during sinus rhythm, with high SE (79%), SP (80%), and accuracy of (79%).[20] This algorithm could be used in patients with CHD as it seems suitable in left and right bundle branch block, premature ventricular contraction, atrial premature beat, and paced beat. Further research on ML ECG interpretation in patients with CHD is warranted and seems feasible.

ML can also be used to make prediction or classification models, which are used to determine in which group a specific outcome would fit. In patients with acquired heart disease the prediction models are mostly used to determine an outcome after surgery or to make a definitive risk model; however, these models perform poorly in predicting outcomes.[45,46] The model by Ruiz et al.[14] in 2016 gave a low accuracy when classifying different diagnoses of congenital heart defects based on a questionnaire. The model made by Ruiz-Fernandez et al.[8] in 2019 could be used in clinical practice because of the high accuracy, but no SE or SP is given. If the SE, SP, or accuracy is low, more research must be carried out or alternative endpoints must be chosen. A solution could be found by comparing the results to human analysis; if it is better than the current gold standard, it could be implemented in clinical practice. However, an accuracy, SE, and SP above 95% should be pursued before implementation is preferred.

The use of AI algorithms in cardiology has gained enormous interest in recent years and is predicted to grow even more in the upcoming years. In patients with acquired heart disease, ML and DL is already being used for imaging modalities and outcome prediction. In patients with CHD, on the other hand, the authors found only 18 articles on learning algorithms. These algorithms have high potential in the population of patients with CHD. Cardiac evaluation in the hospital with ECG and imaging techniques are frequent and a great amount of data is generated from these patients. Moreover, patients with CHD are vulnerable to cardiac morbidity and mortality and often experience complications. Prediction of deterioration in these patients could save lives.

There are a lot of different AI algorithms, with none being superior over the others for their specific task. The authors found that most techniques gave comparable SE, SP, and accuracy. It is more important to choose the right patient features or input data and perform preprocessing of the data to be used.[47-49] Because a large number of different algorithms were used, it was difficult to compare them. It is important avoid deviation in the different algorithms to keep track of the focus: implementation in the clinical setting. In an ideal situation, the authors would like to have an algorithm that would be able to calculate the same values (in case of calculating ventricle volumes) as an imaging physician would. Bigger and more diverse datasets are needed to train the algorithms better and to be applicable to every patient in clinic. A limitation of current algorithms is that even though the accuracy, SE, and SP can be high, mistakes can still occur. This raises questions regarding the extent to which physicians should trust the algorithm and what safeguards are in place if the conclusion of the algorithm is incorrect. However, these questions are beyond the scope of the current review.

CONCLUSION

AI algorithms are increasingly applied in healthcare and draw a lot of attention given the large potential the technology promises. Results of recent studies on AI algorithms in patients with CHD indeed show promising results, as the algorithms aid analysis of ECG, cardiac imaging, and helps to predict outcomes. However, current data on AI algorithms in patients with CHD is still limited and larger scale studies are warranted to provide algorithms that could assist physicians better in the future with high SE, SP, and accuracy.

References

1. van der Bom T et al. The changing epidemiology of congenital heart disease. Nat Rev Cardiol. 2011;8(1):50-60.
2. van der Bom T et al. The prevalence of adult congenital heart disease, results from a systematic review and evidence based calculation. Am Heart J. 2012;164(4):568-75.
3. Schwerzmann M et al. Challenges of congenital heart disease in grown-up patients. Swiss Med Wkly. 2017;147:w14495.
4. Diller GP et al. Utility of machine learning algorithms in assessing patients with a systemic right ventricle. Eur Heart J Cardiovasc Imaging. 2019;20(8):925-31.
5. Schultz KE et al. Extended cardiac ambulatory rhythm monitoring in adults with congenital heart disease: arrhythmia detection and impact of extended monitoring. Congenit Heart Dis. 2019;14(3):410-8.
6. Egbe AC et al. Role of QRS fragmentation for risk stratification in adults with tetralogy of Fallot. J Am Heart Assoc. 2018;7(24):e010274.
7. Saleh A et al. Predictive value of P-wave and QT interval dispersion in children with congenital heart disease and pulmonary arterial hypertension for the occurrence of arrhythmias. J Saudi Heart Assoc. 2019;31(2):57-63.
8. Ruiz-Fernández D et al. Aid decision algorithms to estimate the risk in congenital heart surgery. Comput Methods Programs Biomed. 2016;126(C):118-27.
9. Gharehbaghi A et al. A decision support system for cardiac disease diagnosis based on machine learning methods. Stud Health Technol Inform. 2017;235:43-7.
10. Hannun AY et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med. 2019;25(1):65-9.
11. Faust O et al. Deep learning for healthcare applications based on physiological signals: a review. Comput Methods Programs Biomed. 2018;161:1-13.
12. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017;69S:S36-40.
13. Yang S et al. Use of an artificial neural network to differentiate between ECGs with IRBBB patterns of atrial septal defect and healthy subjects. Med Inform Internet Med. 2002:27(1):49-58.
14. Ruiz VM et al. Early prediction of critical events for infants with single-ventricle physiology in critical care using routinely collected data. J Thorac Cardiovasc Surg. 2019;158(1):234-43.e3.
15. Diller GP et al. Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10 019 patients. Eur Heart J. 2019;40(13):1069-77.
16. Pereira F et al. Automated detection of coarctation of aorta in neonates from two-dimensional echocardiograms. J Med Imaging. 2017;4(1):014502.
17. Haseena HH et al. Classification of arrhythmia using hybrid networks. J Med Syst. 2011;35(6):1617-30.
18. Kim J et al. Algorithm for classifying arrhythmia using extreme learning machine and principal component analysis. Conf Proc IEEE Eng Med Biol Soc. 2007;2007:3257-60. doi:10.1109/IEMBS.2007.4353024.
19. Yang W et al. Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine. Comput Biol Med. 2018;101:22-32.
20. Attia ZI 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):861-7.
21. Elgendi M et al. The unique heart sound signature of children with pulmonary artery hypertension. Pulm Circ. 2015;5(4):631-9.
22. Gharehbaghi A et al. A novel method for screening children with isolated bicuspid aortic valve. Cardiovasc Eng Technol. 2015;6:546-56.
23. Elgendi M et al. The voice of the heart: vowel-like sound in pulmonary artery hypertension. Diseases. 2018;6(2):26.

24. DeGroff CG et al. Artificial neural network–based method of screening heart murmurs in children. Circulation. 2001;103(22):2711-6.
25. Sepehri AA et al. A novel method for pediatric heart sound segmentation without using the ECG. Comput Methods Programs Biomed. 2010;99(1):43-8.
26. Thompson WR et al. Artificial intelligence-assisted auscultation of heart murmurs: validation by virtual clinical trial. Pediatr Cardiol. 2019;40(3):623-9.
27. Bhatikar SR et al. A classifier based on the artificial neural network approach for cardiologic auscultation in pediatrics. Artif Intell Med. 2005;33(3):251-60.
28. Sepehri AA et al. An intelligent phonocardiography for automated screening of pediatric heart diseases. J Med Syst. 2016;40(1):16.
29. Neukamm C et al. Right ventricular volumes assessed by echocardiographic three-dimensional knowledge-based reconstruction compared with magnetic resonance imaging in a clinical setting. Congenit Heart Dis. 2014;9(4):333-42.
30. Nyns ECA et al. Evaluation of knowledge-based reconstruction for magnetic resonance volumetry of the right ventricle after arterial switch operation for dextro-transposition of the great arteries. Int J Cardiovasc Imaging. 2016;32(9):1415-23.
31. Chiogna M et al. An empirical comparison of expert-derived and data-derived classification trees. Stat Med. 1996;15(2):157-69.
32. Folland ED et al. Assessment of left ventricular ejection fraction and volumes by real-time, two-dimensional echocardiography. A comparison of cineangiographic and radionuclide techniques. Circulation. 1979;60(4):760-6.
33. Chowdhury DR et al. An artificial neural network model for neonatal disease diagnosis. Int J Artif Intell Expert Syst. 2011;2(3):96-106.
34. Uğuz H. A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases. J Med Syst. 2012;36(1):61-72.
35. Ari S et al. Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier. Expert Syst Appl. 2010;37(12);8019-26.
36. Asch FM et al. Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert. Circ Cardiovasc Imaging. 2019;12(9):e009303.
37. Genovese D et al. Machine learning-based three-dimensional echocardiographic quantification of right ventricular size and function: validation against cardiac magnetic resonance. J Am Soc Echocardiogr. 2019;32(8):969-77.
38. Ruijsink B et al. Fully automated, quality-controlled cardiac analysis from CMR: validation and large-scale application to characterize cardiac function. JACC Cardiovasc Imaging. 2020;13(3):684-95.
39. Özbay Y et al. A fuzzy clustering neural network architecture for classification of ECG arrhythmias. Comput Biol Med. 2006;36(4):376-88.
40. Rodriguez-Sotelo JL et al. Unsupervised feature selection in cardiac arrhythmias analysis. Poster FrBPo01.27. Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 3-6 September, 2009.
41. Cubanski D et al. A neural network system for detection of atrial fibrillation in ambulatory electrocardiograms. J Cardiovasc Electrophysiol. 1994;5(7):602-8.
42. Özbay Y. A new approach to detection of ECG arrhythmias: complex discrete wavelet transform based complex valued artificial neural network. J Med Syst. 2009;33(6):435-45.
43. Chakroborty S. Accurate arrhythmia classification using auto-associative neural network. Conf Proc IEEE Eng Med Biol Soc. 2013;2016:4247-50. doi:10.1109/EMBC.2013.6610483.
44. Tsipouras MG et al. An arrhythmia classification system based on the RR-interval signal. Artif Intell Med. 2005;33(3):237-50.
45. Lopes RR et al. Value of machine learning in predicting TAVI outcomes. Neth Heart J. 2019;27(9):443-50.
46. Maeno Y et al. A highly predictive risk model for pacemaker implantation after TAVR. JACC Cardiovasc Imaging. 2017;10(10 Pt A):1139-47.
47. Kaya Y et al. Effective ECG beat classification using higher order statistic features and genetic feature selection. Biomed Res. 2017;28(17):7594-603.
48. Zhu J et al. Feature extraction from a novel ECG model for arrhythmia diagnosis. Biomed Mater Eng. 2014;24(6):2883-91.
49. Asgari S et al. Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine. Comput Biol Med. 2015;60:132-42.