



Decision Support Systems in HF based on Deep Learning Technologies

Marco Penso^{1,2} · Sarah Solbiati^{1,3} · Sara Moccia⁴ · Enrico G. Caiani^{1,3}

Accepted: 20 January 2022 / Published online: 10 February 2022
© The Author(s) 2022

Abstract

Purpose of Review Application of deep learning (DL) is growing in the last years, especially in the healthcare domain. This review presents the current state of DL techniques applied to electronic health record structured data, physiological signals, and imaging modalities for the management of heart failure (HF), focusing in particular on diagnosis, prognosis, and re-hospitalization risk, to explore the level of maturity of DL in this field.

Recent Findings DL allows a better integration of different data sources to distillate more accurate outcomes in HF patients, thus resulting in better performance when compared to conventional evaluation methods. While applications in image and signal processing for HF diagnosis have reached very high performance, the application of DL to electronic health records and its multisource data for prediction could still be improved, despite the already promising results.

Summary Embracing the current big data era, DL can improve performance compared to conventional techniques and machine learning approaches. DL algorithms have potential to provide more efficient care and improve outcomes of HF patients, although further investigations are needed to overcome current limitations, including results generalizability and transparency and explicability of the evidences supporting the process.

Keywords Deep learning · Heart failure · Artificial intelligence · Prognosis · Diagnosis · Readmission

Introduction

Heart failure (HF) represents a severe condition affecting approximately 2% of the adult worldwide population, thus counting around 36 million of individuals globally; it consists of a chronic and progressive syndrome characterized by structural or functional cardiac dysfunctions with reduced (HFrEF; < 40%) or preserved (HFpEF; ≥ 50%) left ventricular ejection fraction [1, 2, 3, 4]. HF represents the

most rapidly growing cardiovascular disorder globally. Its pathological spectrum involves numerous symptoms able to greatly affect the patient's quality of life, as dyspnea, fatigue, and poor exercise tolerance, leading to frequent hospitalizations and shortened life expectancy [2, 3].

Cardiac conditions and causes of death vary in the HF population, and although the main underlying causes of this syndrome have been identified, including coronary artery disease, valvular heart disease, hypertension, cardiomyopathies and other (Fig. 1), the prevalence of HF is expected to increase, accounting for a substantial burden to the healthcare system [4, 5]: specifically, due to the ageing population, treatment costs relevant to HF are expected to double by 2030 [6].

Despite advancements in clinical management, surgical procedures, and medical devices in the treatment of all causes associated with HF, significant challenges still persist in current treatments [5], and HF remains one of the main global health concerns [6]. Modeling the driving factors of HF for achieving high prediction accuracy in both diagnosis and prognosis is still an unmet medical need. Accordingly, there is the need for novel approaches to optimize the management of this chronic disease, to improve clinical

This article is part of the Topical Collection on *Digital Medicine in Heart Failure*

✉ Enrico G. Caiani
enrico.caiani@polimi.it

¹ Department of Electronics, Information and Biomedical Engineering, Politecnico Di Milano, P.zza L. da Vinci 32, 20133 Milan, Italy

² Centro Cardiologico Monzino IRCCS, Milan, Italy

³ Institute of Electronics, Information Engineering and Telecommunications (IEIIT), Italian National Research Council (CNR), Milan, Italy

⁴ The BioRobotics Institute, Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna, Pisa, Italy

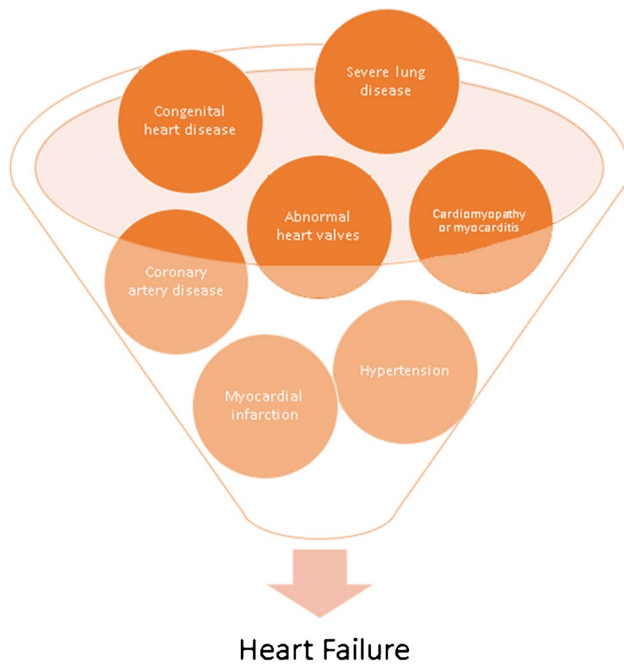


Fig. 1 Schematization of the different pathological conditions that may lead to heart failure

decision-making, and to ultimately reduce related healthcare expenditures. As HF has been recognized as a heterogeneous multifactorial syndrome, improvement in the assessment and management of HF patients requires to integrate data obtained by different sources (e.g., laboratory, echocardiographic and morphologic data) and to handle the complex interplay of various symptoms and comorbidities (both cardiovascular and non-cardiovascular) involved in the HF pathology.

Healthcare is undergoing a new era characterized by the availability of a massive amount of biomedical data, which necessarily opens to new opportunities. The advancement of big data solutions within the healthcare system has allowed to store and manage huge amount of data with the aim to develop new disease risk assessment tools and prediction models, but exploiting in clinical practice these advances leads to unprecedented challenges regarding data analysis and interpretation, as well as many difficulties related to heterogeneity, quality, and integrity of the healthcare data [7].

Machine learning (ML) and deep learning (DL) methods, as a branch of artificial intelligence (Fig. 2), have experienced a rapid growth over the past few years achieving state-of-the-art performance in various domains, including medical imaging, diagnosis, and prognosis [8, 9, 10]. DL is a subfield of ML and represents a family of algorithms that can be used to learn complex and highly predictive patterns that generally remain unexplored using conventional statistic approaches. In contrast to ML, learning solutions based on DL do not require to design a

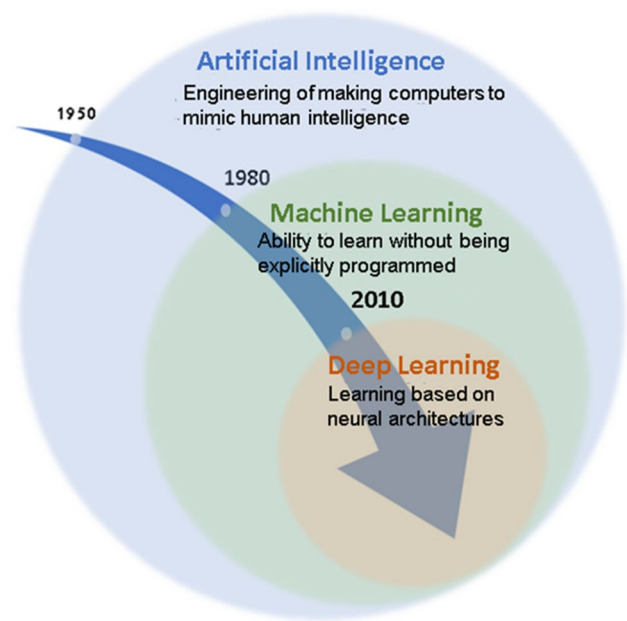


Fig. 2 Evolution of artificial intelligence and its main components, in which deep learning represents a subset of machine learning methods

priori feature extractors from which the learning algorithm detects patterns [11]. In this way, the algorithm is free to learn by itself, automatically defining the features to be considered and the patterns to be searched in order to perform classification or prediction. An additional advantage for DL solutions is the possibility to integrate different structured and unstructured data types as input, which is particularly relevant considering the typical heterogeneity of the healthcare data [11].

In this context, the aim of this review is to provide key concepts for DL that clinicians need to be familiar, and to give an overview of the current advancements of DL research in several clinical applications for the treatment of HF patients.

The paper is organized as follows: in the “[Deep Learning: Key Concepts for Clinicians](#)” section, the fundamental concepts of DL along with a presentation of common DL models used in cardiology are given; in the “[Deep Learning in HF Diagnosis](#)” and “[Deep Learning in HF Prognosis \(End of Hospitalization\)](#)” sections, an overview of recent DL approaches for HF diagnosis and prognosis, respectively, are described. Thereafter, in the “[Deep Learning for Predicting HF Readmission: from EHR to Home Monitoring](#)” section, a review of DL solutions for predicting HF hospital readmission is presented, and in the “[Challenges for Deep Learning](#)” section, current challenges for DL solutions are discussed. Finally, in the “[Conclusion](#)” section, conclusions are drawn with a focus on the future directions for DL applied to the treatment of HF patients.

Deep Learning: Key Concepts for Clinicians

As previously introduced, DL constitutes a subset of ML methods building on the foundations of neural networks, thus trying to mirror the way the human brain processes data (in particular, its learning ability). In contrast to conventional ML methods, requiring human interaction to define which features in the available data have to be considered important for the solution of the classification/prediction problem, DL is meant to learn by itself how to extract knowledge from the data, without being explicitly programmed [12]. In other words, DL automatically extracts the features from the data that are considered important to solve the given task, thus removing the need to select them beforehand, and without involving prior knowledge to explain the observed variability in the data. These fundamental characteristics have generated enthusiasm about the potential of DL to solve problems beyond the human capability, and in recent years, the utilization of DL approaches, that have demonstrated to perform at human-level efficiency and in certain tasks even with higher performance than expert clinicians [12, 13, 14], has surpassed that of ML. The main reasons behind this success, compared to learning algorithms based on hand-designed methods, have to be found in the increasing available computational power, in the larger availability of data, and in the rapid algorithms' development. Indeed, in few years, DL has been able to show its potential in defining new opportunities for improving therapies and treatment, in performing early diagnosis and in reducing the length of hospitalization.

Within DL, artificial neural network (ANN) is an information-processing system, whose structure and functionality simulate the nervous system and the human brain [15]. The main element is the neuron, a simple processing unit, that sends information to other neurons through action potentials, and working in parallel these neurons define a layer. As the brain processes information through multiple stages of transformation, similarly the ANN is characterized by multiple layers of neurons, in order to achieve learning capability [15]. Thanks to the fact that each layer performs a nonlinear mapping based on the previous layer's output, this allows the network to learn via progressive levels of information abstraction [15, 16]. This ability to learn features at multiple level of abstraction allows the network to learning complex functions, without depending on manually developed features.

The most common type of ANN is the convolutional neural network (CNN), which was inspired by the structure of the human visual system. A CNN can be considered an ANN with many identical copies of neurons in its layers, thus utilizing the local relationship within the

data to extract spatial features. This allows the network to increase the number of neurons, and hence its computational power, while keeping the number of learnable parameters relatively small. CNNs are designed to process arrays of data: in 1D as signals (e.g., electrocardiographic, audio, or textual data); in 2D as images; and in 3D as video or volumetric data [16]. Considering an image as input, the first layers of the CNN are associated in learning how to recognize basic lines and curves; moving more deeply, the following layers apprehend shapes and blobs, while in the last layers the ability to classify increasingly complex objects within the image is reached. One of the most popular CNN architecture for medical image analysis is represented by the U-Net [17], that has shown impressive performance, even with a scarce amount of training samples.

Recurrent neural network (RNN) represents another class of ANN, able to recognize patterns in temporal or sequential data [15, 18]. In contrast to common ANN, where the inputs are independent from each other, the main characteristic of RNN is the ability to remember information from prior inputs to generate the current output. In this way, the output of RNN depends on the current input and on all the previous elements of the sequence, where each neuron acts as a memory cell while computing operations. While CNNs are suitable for handling spatial information, RNNs are more suitable for handling temporal or sequential information. For example, given a sequence of frames, a RNN takes the first frame and makes a prediction; the prediction of the following frame is conditioned by the information obtained on the previous frame. Two popular architectures in the RNN family are the gated recurrent units (GRU) and the long short-term memory (LSTM), designed to process information over extended time [19, 21].

More recently, generative adversarial network (GAN) has been introduced in the field of DL, and has increasingly been used in several medical image analyses applications, such as denoizing, reconstruction, segmentation, synthetization, classification, and image-to-image translation. Thanks to its impressive performance, GAN has gained a lot of attention: as its name suggests, unlike conventional ANN, GAN consists of two networks, known as generator and discriminator, trained in an adversarial way [22]. While the generator tries to generate new data, the discriminator learns to distinguish the synthetic data from the real ones. The goal of the discriminator is to force the generator in improving its performance in learning to generate a more realistic data distribution, with the aim of deceiving the discriminator.

For all these methods, the common development procedure starts from the availability of a labeled (i.e., gold standard) dataset, where each data is classified by the expert binary (i.e., healthy and pathologic) or multi-class

classification. This dataset is divided into training, validation, and testing sub-datasets: the training dataset is used to automatically generate the features able to reach the expected goal, compared to the gold standard labels in terms of specificity, sensitivity, and accuracy, often summarized in the area under the receiver operating characteristics (ROC) curve. The validation is used to further tune other parameters in the network in the attempt to further optimize its performance. Finally, the real performance is computed by testing the developed network on the testing dataset.

Deep Learning in HF Diagnosis

An early diagnosis of HF may reduce patients' mortality and morbidity. Consequently, wide efforts have been put in the research to develop algorithms to support clinicians in early diagnosing HF. HF diagnosis may be achieved through the analysis of electrocardiography (ECG), as well as medical images, mainly acquired through magnetic resonance (MRI) and ultrasound (US) images. Furthermore, electronic health records (EHRs, also known as medical records) can be used to this purpose. Early approaches mainly exploited model-based algorithms, while more recently data-driven algorithms (i.e., ML and DL) have shown interesting results given their ability to tackle the complexity and variability of clinical data acquired from subjects at risk of developing HF. The papers surveyed in this section are summarized in Table 1, where the publication date, data source, aim, algorithm, and dataset size are specified.

The widest literature in the field can be found for the processing of the ECG signal. In Kwon et al. [23••], a ML algorithm based on ANN was proposed for HF identification. The ANN processes both demographic and ECG features, achieving an area under the receiver operating characteristics curve (AUC) of 0.89. In Çınar et al. [24•], a more advanced algorithm based on CNNs was used to automatically extract features from the ECG spectrogram. The features were classified using support vector machines (SVMs), achieving an accuracy of 0.97.

A fully DL-based pipeline was proposed by Acharya et al. [25] which exploits a CNN to automatically extract relevant features from the ECG signals and performs an early diagnosis of HF. The pipeline allows to perform end-to-end training, lowering the training time, with an achieved accuracy of 0.99. Lih et al. [26] coupled CNNs with long short-term memory (LSTM) to keep into account the temporal information naturally encoded into the ECG, obtaining an accuracy of 0.98. A similar approach was used in [27], which further included an inception module in the CNN to allow multi-scale analysis, thus achieving an accuracy of 0.99.

A more complex CNN architecture, based on U-Net, was proposed in [28], where residual blocks were exploited to

perform a more accurate feature extraction and classification, reaching an AUC of 0.90. Residual block adds the output of a previous layer to the output of the following layer to extract some additional spatial information. In [29], the first layer of a custom CNN was replaced by Gabor filters to lower the training complexity while extracting relevant high-frequency ECG features, with a reported accuracy of 0.99. Gabor filters are linear filters used for texture analysis and feature extraction, which have been shown excellent localization properties both in spatial and frequency domain, simulating the receptive fields of the human visual system [30••].

As regards other signals, a recent work [29] investigated the possibility to diagnose HF from heart sounds, where logistics regression and gated recurrent units were used to identify the presence of HF: despite the promising results (accuracy = 0.99), more research is still required in this field.

In the last decades, also the analysis of EHRs to perform HF diagnosis has been receiving attention, thanks to the large availability of digitalized data, as well as to the development of more and more accurate ML/DL algorithms. Choi et al. [30••] proposed a milestone paper on the use of recurrent networks for early detecting HF onset (achieving an AUC of 0.88) and, based on it, several works have been published following a similar paradigm. Examples include [31], that used LSTM to process time-stamped EHRs containing medicinal information achieving an AUC of 0.89, and [32], that classified a large variety of features (e.g., demographic, procedural, medicinal features) with LSTM achieving an AUC of 0.82. A more advanced approach was proposed by Ma et al. [33•] that built an embedding from the EHR using CNNs and attention mechanisms, where the embedding was classified with a custom-built prediction model achieving an accuracy of 0.91.

With the goal of predicting HF onset, applications of ML and DL to the field of image processing have been also proposed. In particular, several papers have focused on MRI imaging: in [34], a DL algorithm for the automatic segmentation of the left ventricle as a prior to evaluate the cardiac function in HF patients was proposed, where a Dice similarity coefficient of 0.97 was achieved. A ML method based on k-nearest neighbors was used in [35] to perform texture analysis of myocardial maps and identify early symptoms of HF, achieving an AUC of 0.85.

Besides MRI, other works focused on echocardiographic imaging: Tabassian et al. [36•] analyzed spatiotemporal patterns of echocardiographic deformation curves using k-nearest neighbors with an accuracy of 0.89, while Cikes et al. [37••] evaluated echocardiographic patterns using k-means clustering to identify pathogroups in patients with HF. An interesting attempt of predicting HF markers from chest radiographs with DL was performed by Seah et al. [38] obtaining promising results (AUC = 0.82), but more research

Table 1 Summary of recent studies exploiting deep learning algorithms for diagnosis of HF

Author	Year	Outcome	Data source	Dataset	Algorithm	Results
Kwon et al. [21]	2019	HF (reduced or mid-range to reduced EF) identification	ECG (RR interval recordings)	55,163 ECG from 22,765 patients	ANNs	AUC=0.889 for HF with reduced EF AUC=0.850 for HF with mid-range to reduced EF Accuracy=0.97
Çinar et al. [22]	2021	HF diagnosis	ECG (RR interval recordings)	162 subjects	CNNs and support vector machines	Accuracy = 0.99 Specificity = 0.99 Sensitivity = 0.99
Acharya et al. [23••]	2019	Congestive HF diagnosis	ECG (RR interval recordings)	140,000 ECG segments	CNNs	Accuracy = 0.98 Specificity = 0.98 Sensitivity = 0.99 Positive predictive value = 0.97
Lih et al. [24•]	2020	Congestive HF diagnosis	ECG (RR interval recordings)	262 subjects	CNNs and long short-term memory	Heart rate variability accuracy on BIDMC, NSR, and FD = 0.99 Accuracy on CHF-RR and NSR-RR = 0.88
Wang et al. [25]	2019	Congestive HF diagnosis	ECG (RR interval recordings)	Five open datasets (BIDMC, NSR, FD, NSR-RR, and CHF-RR) for a total of 178 subjects	CNNs and long short-term memory	AUC = 0.90 Accuracy = 0.89 Accuracy = 0.99
Lei et al. [26]	2021	Congestive HF diagnosis	ECG (RR interval recordings)	83 subjects	CNNs based on U-Net	AUC = 0.88 (18-month observation period)
Jahmunah et al. [27]	2021	Congestive HF diagnosis	Lead II ECG signals	262 subjects (92 healthy controls, 7 CAD, 148 MI, and 15 CHF)	CNNs and Gabor filtering	AUC = 0.89
Choi et al. [30••]	2017	HF diagnosis	EHRs	3884 HF + 28,903 control subjects	RNNs with gated recurrent units	AUC = 0.82
Maragatham et al. [31]	2019	HF diagnosis	EHRs	4289 HF + 30,249 control subjects	RNNs with long short-term memory	Accuracy = 0.91
Rasmy et al. [32]	2018	Prediction of HF risk	EHRs	150,000 HF + 1,000,000 control subjects	RNN with long short-term memory	Accuracy = 0.99
Ma et al. [33•]	2019	HF diagnosis	EHRs	4925 HF	CNNs	DSC = 0.97
Gao et al. [29]	2020	HF diagnosis	Heart sounds	2543 subjects (42 HF reduced EF, 66 HF preserved EF, 2435 controls)	Gated recurrent unit model	AUC = 0.85
Lan et al. [34]	2019	Left ventricle segmentation for EF evaluation	MRI	45 volumes from 45 subjects (12 HF with ischemia, 12 HF without ischemia, 12 hypertrophy, 9 normal)	CNNs	Accuracy = 0.85 AUC = 0.89
Baessler et al. [35]	2019	Diagnosis of HF myocarditis	MRI	71 subjects	k-Nearest neighbors	Four phenogroups were identified. Group 1 and 3 reported a 64% and 65% reduction in the risk of HF or death, respectively (hazard ratio 0.36, $p=0.001$) after defibrillator treatment
Tabassian et al. [36•]	2018	Diagnosis of HF	Echocardiography stress test (strain parameters)	100 subjects (33 with HF preserved EF, 67 controls)	k-Nearest neighbors	AUC = 0.82
Cikes et al. [37••]	2019	Phenogrouping a HF cohort	Echocardiographic data and clinical parameters	1106 HF subjects	Unsupervised ML algorithm (k-means clustering)	
Seah et al. [38]	2019	Diagnosis of congestive HF	Radiography	103,489 frontal chest radiographs in 46,712 patients	CNNs	

CAD, coronary artery disease; MI, myocardial infarction; CHF, congestive heart failure; HF, heart failure; HERS, electronic health records; ECG, electrocardiography; MRI, magnetic resonance images; AUC, area under the curve; ANN, artificial neural network; CNNs, convolutional neural networks; EF, ejection fraction; RNNs, recurrent neural networks; DSC, dice similarity coefficient; BIDMC: Beth Israel Deaconess Medical Center CHF database; NSR: normal sinus rhythm; FD: Fantasia database

is still needed to understand the potentiality of DL in processing chest radiographs for HF diagnosis.

Deep Learning in HF Prognosis (End of Hospitalization)

Several studies have used DL to predict different outcomes in HF patients [39, 40•]. Specifically, the measured outcomes that were studied include mortality, hospitalizations, readmissions, risk prediction, need for mechanical circulatory support, heart transplantation, and treatment effect (Table 2). The general process of DL techniques regarding HF prognosis is based on data obtained through the EHRs that might include demographic information, treatment and medication, laboratory results, ECG, and echocardiographic findings before and during hospital stay. Wang et al. [41••] applied ANN for early detection of patients' HF death in three observation windows (i.e., in-hospital, 1-month and 1-year mortality), studying 10,203 in-patient EHRs. It is noteworthy the introduction of a focal loss function [42] into the proposed framework, to deal with the imbalanced class problem, and a feature rearrangement layer to improve feature representation of the convolutional network. The proposed ANN provided an AUC in predicting mortality of 0.904 (in-hospital), 0.891 (1-month observation), and 0.887 (1-year observation).

Kwon et al. [43] used a DL-based model in a multicenter cohort of acute HF patients for predicting in-hospital mortality, and at 12 and 36 months, by integrating clinical and laboratory data. Training included 2165 patients, while validation was performed on 4759, reaching an AUC for predicting in-hospital and 12 and 36 months mortality of 0.880, 0.782, and 0.813, respectively. Overall, DL outperformed both the conventional Get with the Guidelines–Heart Failure (GWTG-HF) score, and the Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC) score [44, 45], as well as other ML models. Since GWTG-HF and MAGGIC cannot be used for initial treatment or screening, Kwon et al. [46•] applied DL to predict in-hospital mortality only on echocardiographic data in 25,776 patients. In a subgroup analyses of HF, DL provided an AUC (0.913) higher than both MAGGIC (0.806) and GWTG-HF (0.783) scores. Medved et al. [47] compared the International Heart Transplantation Survival Algorithm (IHTSA) based on DL, with the Index for Mortality Prediction After Cardiac Transplantation (IMPACT), for predicting 1-year survival after heart transplantation. In 27,705 patients (5597 in the test cohort), DL exhibited an AUC of 0.654, with improved performance compared to the IMPACT model (AUC 0.608). Although IHTSA was designed to predict long-term survival, it showed better discrimination at 1-year mortality than IMPACT.

Therefore, even though modest, these results are promising for DL techniques applications in clinical practice.

To model early HF readmission prediction, a deep unified network, an innovative architecture designed to avoid overfitting including both structured (i.e., demographics, clinical and laboratories results) and unstructured (physician notes and discharge summaries) data from EHRs of 11,510 patients, was applied [48•]. Obtaining an AUC of 0.705, the developed 30-day readmission model reported the best performance compared to logistic regression (LR) (0.664), gradient boosting (0.650), and maxout networks (0.695). In a novel study applying DL to EHRs for treatment effect prediction on 736 HF patients, the proposed generated GAN learning strategy outperformed benchmark models in terms of both accuracy (0.688) and AUC (0.654) [49•]. The DL treatment effect prediction model used two auto-encoders for learning features of both patient characteristics and treatments from EHRs. Specifically, the DL scheme could generate and discriminate the predicted treatments from the real ones so that highly representative features were extracted from the EHRs data [49•].

In [50•], 93,260 HF patients were analyzed to identify preventable outcomes, such as hospitalization and emergency department visits. Compared to ML and LR models, DL produced the highest AUC of 0.778 and 0.681, respectively. Remarkable was the effort of Li et al. [51••] in terms of interpreting DL models, by developing an interactive clinical risk prediction system based on RNN with an intuitive visualization design, increasing transparency to the information infrastructure, thus allowing visual interpretation of the prediction results. On 554 HF and 1662 control patients, the proposed DL model outperformed the state-of-the-art approaches by approximately 1.5%. Recently, Lu et al. [52••] proposed a DL approach to model long-term and short-term HF clinical trajectories on 8093 patient with congenital heart disease. The network outperformed various baseline models and was able to predict different types of patient trajectories (AUC 0.863). A separate study used DL to add prognostic value of data acquired from a cardiopulmonary exercise test (CPET) [53]. In another study involving 1156 HF patients, DL demonstrated little improvement compared to statistical model (AUC 0.842 vs. 0.837), while both were superior to CPET-risk score (AUC 0.759) [54].

Finally, ML and DL seem promising in identifying distinct patient subgroups with HFpEF using unsupervised learning to deliver more tailored clinical care. These techniques make it possible to learn from an input dataset without the need for training with labeled data (expected outputs). In fact, the model learns to draw inferences and identify significant features within the unlabeled data space, for the purposes of clustering or data reduction. The pathological development of HFpEF has been attributed to a complex interplay of cardiac and extracardiac dysfunctions [55, 56] leading to a marked phenotypic heterogeneity among patients of this population. This diversity highlights the fact that there is not a single

Table 2 Summary of recent studies exploiting deep learning algorithms in HF prognosis

Author	Year	Outcome	Data source	Dataset	Algorithm	Results
Medved et al. [47]	2018	Survival prediction after heart transplantation	EHRs	27,705 patients	ANNs	Reduction of 12% for ROC and 10% for C-index by using deep learning technique
Wang et al. [41••]	2020	Mortality prediction	EHRs	10,203 patients	CNNs	AUC in-hospital 0.904, 1-month 0.891, 1-year 0.887
Golas et al. [48•]	2018	Readmission prediction	EHRs	11,510 patients	Deep unified networks	AUC 0.705
Kwon et al. [43]	2019	Mortality prediction	EHRs	6924 patients	ANNs	AUC in-hospital 0.880, 12-month 0.782, 36-year 0.813
Lewis et al. [50•]	2021	Preventable hospitalizations, emergency department and costs	Clinical history	93,260 patients	ANNs	AUC for deep learning were 0.778, 0.681, and 0.727, respectively
Ashfaq et al. [39]	2019	Readmission prediction	EHRs	7655 patients	RNNs with long short-term memory	AUC 0.77
Chu et al. [49•]	2020	Treatment effect prediction	EHRs	736 patients	GAN	AUC 0.688
Kwon et al. [46•]	2019	In-hospital mortality	Clinical + echocardiography	760 HF	ANNs	AUC 0.913
Li et al. [51••]	2020	Risk prediction	EHRs	554 HF + 1662 controls	RNNs	RNN outperforms the state-of-the-art approaches by approximately 1.5%
Pandey et al. [59]	2021	Phenotyping diastolic dysfunction in HF with preserved ejection fraction	Echocardiography	1242 patients	ANNs	AUC 0.88
Hearn et al. [53]	2018	Clinical deterioration	Cardiopulmonary exercise test data	1156 HF	ANNs	AUC 0.842
Lu et al. [52••]	2021	Long-term trajectory prediction	EHRs	8093 HF	RNNs with gated recurrent units	AUC 0.863

HF, heart failure; *EHRs*, electronic health records; *AUC*, area under the curve; *ANN*, artificial neural network; *CNNs*, convolutional neural networks; *RNNs*, recurrent neural networks; *GAN*, generative adversarial networks

pathological process underlying the observed dysfunction, thus affecting the targeted management plan. Recently, different studies made progress in clustering HFpEF patients by integrating multiple patient data, as step towards personalizing treatment and improving prognosis of the disease [57, 58]. Pandey et al. [59] analyzed 1242 HFpEF to predict high- and low-risk phenogroups and validated the network in 5 external cohorts. The DL approach showed higher AUC than the 2016 American Society of Echocardiography guideline-based left ventricular grades [60] for predicting elevated left ventricular filling pressure (0.883 vs. 0.676). Kaptein et al. [61] proposed an unsupervised learning approach to identify subgroups of patients with asymptomatic diastolic dysfunction, where three subgroups were identified. Similarly, in [62], a model-based clustering on clinical and echocardiogram variables in 320 HFpEF patients was applied, from which six phenogroups

were derived. Although HFpEF remains a challenging clinical condition to manage, clustering patients with model-based learning using echocardiographic and EHRs data may provide better granularity with improved prognostic benefit for patients with HFpEF compared to the current clinical paradigm, thus creating phenotype clusters that are strongly linked to survival. This new approach may lead to improved personalized care pathways for treating patients with HF.

Deep Learning for Predicting HF Readmission: from EHR to Home Monitoring

The significantly high rate of readmissions in hospital after HF, with 61.3% of the patients being readmitted for HF within 1 year after discharge [63], has a negative impact

on patients' quality of life, as well as on the healthcare systems. Therefore, it appears crucial to develop efficient tools in order to predict patient's re-hospitalization probability and related causes of readmission. This would primarily help tailor patients' remote support and education after discharge. Also, the early identification of patients at higher risk would improve the scheduling of potentially life-saving follow-ups. Accordingly, several works focused on this problem, exploring the use of DL methods applied to different types of data. Those studies are discussed below, and summarized in Table 3.

Since discharge, e-Health solutions could support the prediction of patient's outcome and probability of re-hospitalization. Among these, EHRs contain a huge amount of data, ranging from anthropometrics and demographic to prescribed therapies, comorbidities, and vital signs. Several works in the literature examined the possibility to apply DL models to EHRs in order to make accurate predictions of hospital readmissions in HF patients. An example of the increasing interest towards this field is represented by CONTENT [64•], a DL model based on a RNN with gated recurrent unit aiming at predicting 30-day hospital readmissions. It was developed using the EHRs of 5393 congestive HF patients, embedding data relevant to patients' diseases, laboratory tests, and medications. Although outperforming other existing models, the results obtained in this work remain unsatisfactory, with 38.94% mean precision-recall AUC, 61.03% receiver operating characteristic AUC, and 69.34% accuracy. Similar results were also obtained with a multi-layer perceptron ANN applied to a linked administrative health dataset (10,757 over-65 HF patients) obtained from the Western Australian Data Linkage System [65•], as well as with a RNN combined with conditional random fields applied to a large hospital claims dataset [66•]. A deep unified network model developed on data obtained during inpatient and outpatient visits provided slightly improved results, with mean AUC equal to 70.5% and an accuracy of 76.4% in predicting 30-day readmission in HF patients [48•].

A key characteristic for a successful introduction of an AI model in the clinical practice stands in its interpretability, thus generally resulting in a higher propensity towards ML compared to DL techniques. Attention-based neural network prediction models represent a valid solution. A recent study [40•] evaluated the possibility to predict all-cause readmission in HF patients within 1 year after discharge using an attention-based neural network built on data contained in the EHRs of 736 HF patients. The proposed model assigns to each feature an "attention weight" indicating its importance in predicting readmission, and thus supporting clinicians in identifying patients at higher risk of a forthcoming relapse. For example, the analysis of the levels of B-type natriuretic peptide is widely used in clinical practice for the diagnosis of HF. As expected, this clinical feature was associated with

considerably higher attention rates in the majority of the patients compared to the other features. Results appeared promising, although the achieved statistics, including mean F1-score and AUC values, remained below 80%, and thus requiring further improvements.

The application of interpretable DL methods could also take advantage of big data coming from EHRs in order to characterize subtypes of HF patients. An example can be observed in the study of Xiao and colleagues [64•], in which the authors were able to identify 20 subgroups of congestive HF patients, each possibly exhibiting different comorbidities that could impact the progression of this syndrome and consequently the readmission risk [3]. This approach could pave the way towards the identification and development of personalized and targeted home-care pathways.

In this context, remote monitoring solutions could effectively support HF patients in managing their condition and improving their quality of life, thus reducing the risk of readmission and mortality [67, 68]. Current methods primarily include telemonitoring with implantable devices, such as in the CardioMems [69] and the IN-TIME approach [70], which are recommended as Class II for use in selected patients by the 2016 ESC guidelines for the diagnosis and treatment of acute and chronic HF [3]. Thanks to the advances in technology and communication systems, non-invasive telemedicine solutions, including telephone-based monitoring and education, wearable and mobile health, have been implemented and tested, appearing particularly promising for patients that are not assigned to an implanted monitoring approach.

For example, remote monitoring of body weight is recommended in HF patients. However, daily monitoring of this parameter alone showed no evidence in identifying higher risk patients [71, 72]. A successful example of non-invasive multi-parametric remote patient monitoring is represented by the TIM-HF2 (Telemedical Interventional Management in Heart Failure II) prospective randomized controlled trial, with 796 patients assigned to the remote monitoring group and 775 to the control group [73••]. The study involved the daily measurement and transmission of several physiological parameters, including body weight, blood pressure, electrocardiogram, heart rate, and peripheral capillary oxygen saturation, as well as a self-rated score of the health status, fostering the cooperation of the telemedical center with cardiologists and general practitioners. The collected data were analyzed using the CE-marked Fontane telemedicine software (T-Systems International GmbH, Frankfurt, Germany), which integrates scalable business intelligence methods in order to assign a patient to a risk category [67]. Results showed that this approach effectively supported the identification of higher risk patients, accelerating tailored intervention and consequently reducing days lost during 1 year of follow-up and all-cause mortality. A recent study further improved these results, implementing a DL neural

Table 3 Summary of recent studies exploiting deep learning algorithms for predicting HF readmission after discharge

Author	Year	Outcome	Data source	Dataset	Algorithm	Results
Xiao et al. [64•]	2018	30-day hospital readmission prediction	EHRs	5393 congestive HF patients	RNN with gated recurrent unit	AUC: 61.03% PR-AUC: 38.94% Accuracy: 69.34%
Awan et al. [65•]	2019	30-day readmission prediction or death	Linked administrative health dataset	10,757 over-65 HF patients	Multi-layer perceptron ANN	AUC: 62.8% PR-AUC: 46.1% Accuracy: 64.93% Sensitivity: 48.42% Specificity: 70.01%
Allam et al. [66•]	2019	30-day readmission prediction	Hospital claims dataset	272,778 patients, 343,328 HF admissions	Several deep learning models were tested. RNN combined with conditional random fields outperformed the others	AUC: 64.2%
Golas et al. [48•]	2018	30-day readmission prediction	EMRs	11,510 HF patients	Deep unified network	AUC: 70.5% Accuracy: 76.4%
Chen et al. [40•]	2020	1-year readmission prediction	EHRs	736 heart failure	Attention-based neural network	AUC: 69.1% Accuracy: 66.7% F1: 74.9% Recall: 79.5% Precision: 71%
Koehler et al. [73••]	2018	Compare non-invasive multi-parameter remote monitoring of HF patients with usual care to identify patients at higher risk	Body weight, blood pressure, electrocardiogram, heart rate, peripheral capillary oxygen saturation, self-rated score of the health status	796 patients assigned to the remote monitoring group and 775 to the control group	Fontane software (T-Systems International GmbH, Frankfurt, Germany), integrates business intelligence algorithms	Percentage of days lost: 4.88% in the remote patient management group and 6.64% in the usual care group ($p=0.0460$) All-cause death rate (100 person-years of follow-up): 7.86 in the remote patient management group compared with 11.34 in the usual care group ($p=0.0280$)
Gontarska et al. [74••]	2021	Risk score prediction	Age, weight, blood pressure, oxygen saturation, gender, diabetes, NYHA class, symptoms and signs of heart failure, ECG-extracted heart rate, sinus rhythm, ventricular tachycardia, atrial fibrillation, self-assessed state of health, weight difference, social variables	763 patients	Deep neural network	AUC: 84% PR-AUC: 19%
Stehlik et al. [76••]	2020		Wearable multi-parameter sensor: heart rate and its variability, arrhythmia burden, respiratory rate, physical activity, body posture	100 patients	Similarity-based machine learning algorithms	10-day window: AUC: 85.7% (HF hospitalization), 80.4% (unplanned non-trauma hospitalization) Event-specific: AUC: 89.3% (HF hospitalization), 83.6% (unplanned non-trauma hospitalization)

HF, heart failure; EHRs, electronic health records; EMR, electronic medical records; ECG, electrocardiography; NYHA, New York Heart Association; AUC, area under curve; PR-AUC, area under the precision-recall curve; ANN, artificial neural network; RNN, recurrent neural network

network model based on the TIM-HF2 database, which allowed to reach a mean AUC value of 84% [74••].

Wearable devices could additionally promote the continuous monitoring of patients' health after discharge, thus representing an opportunity to improve remote monitoring and healthcare [75]. The LINK-HF study aimed at evaluating the accuracy of predicting deterioration which leads to re-hospitalization in HF patients using a wearable sensor (Vital Connect, San Jose CA) worn on the chest [76••]. Of note, this device recorded continuous acquisition of ECG, accelerometric signal, skin impedance, and skin temperature, thus permitting the monitoring of heart rate and its variability, arrhythmia burden, respiratory rate, physical activity, and body posture. Collected data were streamed to a smartphone and analyzed in Cloud. Similarity-based ML algorithms were able to generate a multivariate index which indicates the level of change of the acquired vital parameters. The presented platform appeared successful in predicting patient's readmission due to worsening HF with a sensitivity from 76.0 to 87.5% at a specificity level of 85%.

Challenges for Deep Learning

DL has demonstrated promising results with better performance in HF evaluation compared to ML and conventional algorithms, which could not be expected a priori. As DL algorithms require larger amount of data in order to provide high-quality results, this may limit their development especially in a clinical context, considering that the labeling data procedure is a time-consuming and tedious task for expert clinicians. Moreover, normal cases are often predominant over pathological ones, leading to unbalanced datasets which may originate biased predictions.

In contrast to conventional diagnostic and prognostic models, and similarly to ML, DL does not assume linear relationship among variables, leading to a better patient-level therapy treatment decisions. In some clinical trials, DL provided performance comparable to those of statistical linear models as logistic regression, suggesting how, depending on the type of data, a different analysis might be more suitable. Specifically, future studies could facilitate the integration of ML/DL models with statistical classifiers.

In addition, DL application in healthcare poses more challenges because data are often highly heterogeneous, noisy, and incomplete, and the number of available patients is usually limited, thus complicating the proper convergence of the DL algorithm and reliability of the results (i.e., garbage in results in garbage out). Moreover, the repeatability of the performance obtained with supervised models trained on specific datasets (i.e.,

monocentric, or obtained using the same equipment) onto data collected within other centers as well as with other equipment, or with different underlying patient factors (i.e., gender distribution, ethnicity, morbidities), needs to be further validated to avoid introducing biases in the results. Therefore, a standardized framework on how to perform and validate clinical studies would be required before implementation of DL into routine clinical use. Indeed, the impact of DL on the clinical decision-making process, on resources utilization and on value-based practice, has not been yet properly investigated. Moreover, the current literature reports an unbalance distribution of studies between ML and DL, with a limited number of DL studies, probably due to the limited availability of data. Indeed, in [77], the authors suggest that a substantially investment will be required in order to create high-quality annotated datasets for the development and the success of DL methods.

Another main limitation of DL models is inherent to the limited explicability of their results in a way that clinicians could understand. Opposite to ML, as the features are determined by the network itself, without a relation with possible features that a human could extract (i.e., mean, standard deviation, common parameters in the temporal or in the frequency domain), often it is not possible to understand which parameters and why have contributed to the generated output. This is particularly critical for decision support systems, where there is the need for the physician to comprehend and evaluate the source of the suggested action before taking the final decision and associated responsibility. Also in other fields, ethical issues have been raised concerning poor explicability, possibly leading to severe consequences [78, 79]. This condition of non-interpretability collides against the concept of evidence-based medicine, the cornerstone for clinical applications of DL, thus potentially limiting its utilization into clinical practice [80]. Possible solutions to cope with this limitation consist in the introduction of attention-base explainable DL methods, where the network is forced to learn on pre-defined attention maps on the original data, that can be visualized to better understand the origin of its results.

An additional aspect that could limit diffusion of DL in the medical field concerns the need for clinical assessment related to the software certification as medical device, as currently regulated by the EU legislation [81]. In fact, such software must undergo approval by notified bodies before being introduced in clinical practice, and proper accuracy and increased benefit over risk need to be demonstrated a priori. Additionally, as these networks are currently evolving based on the constant availability of data, the problem of re-certification over time has been posed to verify the same longitudinal performance.

Conclusion

It is evident from the literature that DL algorithms have witnessed increasing applications in different aspects of the management of HF patients, with the aim to improve efficiency in diagnosis and prognosis. These methods have already demonstrated to overcome the performance of conventional approaches in different clinical setting, being able to integrate different data sources in order to improve diagnosis and prediction, potentially leading to tailored treatments.

The results described in this review have illustrated the potential capabilities of DL methods to improve prediction relevant to mortality and hospital readmission, highlighting how these promising tools could introduce substantial positive and significant changes in the clinical workflow in the future treatment of HF in the near future. For example, the increasing availability of smart analysis in EHRs based on DL applications will reduce the need for scoring systems, enabling personalized treatment for HF patients. DL analytical skills have been shown to be superior to those of expert clinicians, since humans can handle only a limited number of cognitive information (i.e., variables in structure data) at once [82, 83], thus facilitating clinical support for early HF risk identification. With a rapidly growing scenario in cardiovascular medicine, DL has the potential of paving the way towards a new generation of predictive methods in healthcare that could automatize essential processes involved in treatment planning, helping in identifying hidden information in complex and heterogeneous datasets to effectively support clinicians in their daily activities. In this scenario, DL has showed potential to classify HF patients into novel phenotypes who might benefit of specific treatments, as well as for early diagnosis of HF to improve its prognosis. The integration of different data sources including EHRs, genomics, and remote patient monitoring could provide a better description on the HF patient individual status, which might support clinicians regarding appropriate intervention and therapy, hospital discharge, and hospital re-admissions. However, for DL to become part of clinical practice, several ethical and regulatory issues need to be properly addressed and solved. These challenges introduce both new opportunities and the need of further research to provide more evidence about the effective benefit of these algorithms in being translated into better quality of care for patients, improved outcomes, and lower healthcare costs.

Acknowledgements The editors would like to thank Dr. Jasper Tromp for taking the time to review this paper.

Declarations

Conflict of Interest The authors declare no competing interests.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
- Of outstanding importance

1. Ziaeeian B, Fonarow GC. Epidemiology and aetiology of heart failure. *Nat Rev Cardiol*. 2016;13:368–78. <https://doi.org/10.1038/nrcardio.2016.25>.
2. Arrigo M, Jessup M, Mullens W, Reza N, Shah AM, Sliwa K, et al. Acute heart failure. *Nat Rev Dis Primers*. 2020;6:16. <https://doi.org/10.1038/s41572-020-0151-7>.
3. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J*. 2016;37:2129–200. <https://doi.org/10.1093/eurheartj/ehw128>.
4. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, et al. Heart disease and stroke statistics-2017 update: a report from the American Heart Association. *Circulation*. 2017;135:e146–603. <https://doi.org/10.1161/CIR.0000000000000485>.
5. Ambrosy AP, Fonarow GC, Butler J, Chioncel O, Greene SJ, Vaduganathan M, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. *J Am Coll Cardiol*. 2014;63:1123–33. <https://doi.org/10.1016/j.jacc.2013.11.053>.
6. Zolfaghar K, Meadem N, Teredesai A, Roy SB, Chin SC, Muckian B. Big data solutions for predicting risk-of-readmission for congestive heart failure patients. In *Big Data, 2013 IEEE International Conference on 2013*:64–71. <https://doi.org/10.1109/BigData.2013.6691760>.
7. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial intelligence in cardiology. *J Am Coll Cardiol*. 2018;71:2668–79. <https://doi.org/10.1016/j.jacc.2018.03.521>.
8. Kwon JM, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. *J Am Heart Assoc*. 2018;7: e008678. <https://doi.org/10.1161/JAHA.118.008678>.

9. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318:2211–23. <https://doi.org/10.1001/jama.2017.18152>.
10. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25:24–9. <https://doi.org/10.1038/s41591-018-0316-z>.
11. Krittanawong C, Johnson KW, Rosenson RS, Wang Z, Aydar M, Baber U, et al. Deep learning for cardiovascular medicine: a practical primer. *Eur Heart J*. 2019;40:2058–73. <https://doi.org/10.1093/eurheartj/ehz056>.
12. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316:2402–10. <https://doi.org/10.1001/jama.2016.17216>.
13. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ*. 2020;368: m689. <https://doi.org/10.1136/bmj.m689>.
14. Kruse R, Borgelt C, Klawonn F, Moewes C, Steinbrecher M, Held P. Computational intelligence: a methodological introduction. 1st ed. London Ltd: Springer; 2018.
15. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–44. <https://doi.org/10.1038/nature14539>.
16. O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2015. pp. 234–241.
17. Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med*. 2020;7:25. <https://doi.org/10.3389/fcvm.2020.00025>.
18. Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Conference on Empirical Methods in Natural Language Processing. ACL; 2014. pp. 1724–34.
19. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9:1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>.
20. Lan L, You L, Zhang Z, Fan Z, Zhao W, Zeng N, et al. Generative adversarial networks and its applications in biomedical informatics. *Front Public Health*. 2020;8:164. <https://doi.org/10.3389/fpubh.2020.00164>.
21. Kwon JM, Kim KH, Jeon KH, Kim HM, Kim MJ, Lim SM, et al. Development and validation of deep-learning algorithm for electrocardiography-based heart failure identification. *Korean Circ J*. 2019;49:629–39. <https://doi.org/10.4070/kcj.2018.0446>.
22. Çınar A, Tuncer SA. Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ECG signals using LSTM and hybrid CNN-SVM deep neural networks. *Comput Methods Biomech Biomed Engin*. 2021;24:203–14. <https://doi.org/10.1080/10255842.2020.1821192>.
- 23.●● Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M, et al. Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. *Appl Intell*. 2019;49:16–27. <https://doi.org/10.1007/s10489-018-1179-1>. **Milestone study in the field of congestive heart failure diagnosis from ECG data using deep convolutional neural networks.**
- 24.● Lih OS, Jahmunah V, San TR, Ciaccio EJ, Yamakawa T, Tanabe M, et al. Comprehensive electrocardiographic diagnosis based on deep learning. *Artif Intell Med*. 2020;103: 101789. <https://doi.org/10.1016/j.artmed.2019.101789>. **The work discusses in details deep learning algorithms for performing HF diagnosis from ECG data.**
25. Wang L, Zhou X. Detection of congestive heart failure based on LSTM-based deep network via short-term RR intervals. *Sensors (Basel)*. 2019;19:1502. <https://doi.org/10.3390/s19071502>.
26. Lei M, Li J, Li M, Zou L, Yu H. An improved UNet++ model for congestive heart failure diagnosis using short-term RR intervals. *Diagnostics (Basel)*. 2021;11:534. <https://doi.org/10.3390/diagnostics11030534>.
27. Jahmunah V, Ng EYK, San TR, Acharya UR. Automated detection of coronary artery disease, myocardial infarction and congestive heart failure using GaborCNN model with ECG signals. *Comput Biol Med*. 2021;134: 104457. <https://doi.org/10.1016/j.compbiomed.2021.104457>.
28. Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*. 1996;381:607–9. <https://doi.org/10.1038/381607a0>.
29. Gao S, Zheng Y, Guo X. Gated recurrent unit-based heart sound analysis for heart failure screening. *Biomed Eng Online*. 2020;19:3. <https://doi.org/10.1186/s12938-020-0747-x>.
- 30.●● Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. *J Am Med Inform Assoc*. 2017;24:361–70. <https://doi.org/10.1093/jamia/ocw112>. **Milestone study in the field of heart failure diagnosis from EHRs data using recurrent neural networks.**
- 31.● Maragatham G, Devi S. LSTM model for prediction of heart failure in big data. *J Med Syst*. 2019;43:111. <https://doi.org/10.1007/s10916-019-1243-3>. **The work proposes a robust and accurate machine learning algorithm for predicting HF from EHRs data.**
32. Rasmay L, Wu Y, Wang N, Geng X, Zheng WJ, Wang F, et al. A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set. *J Biomed Inform*. 2018;84:11–6. <https://doi.org/10.1016/j.jbi.2018.06.011>.
33. Ma F, Wang Y, Xiao H, Yuan Y, Chitta R, Zhou J, et al. Incorporating medical code descriptions for diagnosis prediction in healthcare. *BMC Med Inform Decis Mak*. 2019;19(Suppl 6):267. <https://doi.org/10.1186/s12911-019-0961-2>.
34. Lan Y, Jin R. Automatic segmentation of the left ventricle from cardiac MRI using deep learning and double snake model. *IEEE Access*. 2019;7:128641–50. <https://doi.org/10.1109/ACCESS.2019.2939542>.
35. Baessler B, Luecke C, Lurz J, Klingel K, Das A, von Roeder M, et al. Cardiac MRI and texture analysis of myocardial T1 and T2 maps in myocarditis with acute versus chronic symptoms of heart failure. *Radiology*. 2019;292:608–17. <https://doi.org/10.1148/radiol.2019190101>.
- 36.● Tabassian M, Sunderji I, Erdei T, Sanchez-Martinez S, Degiovanni A, Marino P, et al. Diagnosis of heart failure with preserved ejection fraction: machine learning of spatiotemporal variations in left ventricular deformation. *J Am Soc Echocardiogr*. 2018;31:1272–1284.e9. <https://doi.org/10.1016/j.echo.2018.07.013>. **The work investigates the problem of HF diagnosis with machine learning. To this goal, the spatiotemporal changes in ventricle deformation are considered, achieving accurate results.**
- 37.●● Cikes M, Sanchez-Martinez S, Claggett B, Duchateau N, Piella G, Butakoff C, et al. Machine learning-based phenotyping in heart failure to identify responders to cardiac resynchronization therapy. *Eur J Heart Fail*. 2019;21:74–85. <https://doi.org/10.1002/ejhf.1333>. **Interesting work on phenotyping in heart failure from echocardiographic data using machine learning.**
38. Seah JCY, Tang JSN, Kitchen A, Gaillard F, Dixon AF. Chest radiographs in congestive heart failure: visualizing neural

- network learning. *Radiology*. 2019;290:514–22. <https://doi.org/10.1148/radiol.2018180887>.
39. Ashfaq A, Sant'Anna A, Lingman M, Nowaczyk S. Readmission prediction using deep learning on electronic health records. *J Biomed Inform*. 2019;97:103256. <https://doi.org/10.1016/j.jbi.2019.103256>.
 - 40.● Chen P, Dong W, Wang J, Lu X, Kaymak U, Huang Z. Interpretable clinical prediction via attention-based neural network. *BMC Med Inform Decis Mak*. 2020;20:131. <https://doi.org/10.1186/s12911-020-1110-7>. **This study exploits Electronic Health Records' data to predict 1-year readmission of HF patients using an interpretable, attention-based neural network model.**
 - 41.●● Wang Z, Zhu Y, Li D, Yin Y, Zhang J. Feature rearrangement based deep learning system for predicting heart failure mortality. *Comput Methods Programs Biomed*. 2020;191: 105383. <https://doi.org/10.1016/j.cmpb.2020.105383>. **This study proposed a framework for handling the class imbalance problem and a feature rearrangement for improving the information flow through the convolutional layers.**
 42. Mulyanto M, Faisal M, Prakosa SW, Leu JS. Effectiveness of focal loss for minority classification in network intrusion detection systems. *Symmetry*. 2021;13:4. <https://doi.org/10.3390/sym13010004>.
 43. Kwon JM, Kim KH, Jeon KH, Lee SE, Lee HY, Cho HJ, et al. Artificial intelligence algorithm for predicting mortality of patients with acute heart failure. *PLoS ONE*. 2019;14: e0219302. <https://doi.org/10.1371/journal.pone.0219302>.
 44. Lagu T, Pekow PS, Shieh M-S, Stefan M, Pack QR, Kashef MA, et al. Validation and comparison of seven mortality prediction models for hospitalized patients with acute decompensated heart failure. *Circ Hear Fail*. 2016;9. <https://doi.org/10.1161/CIRCHEARTFAILURE.115.002912>.
 45. Pocock SJ, Ariti CA, McMurray JJV, Maggioni A, Køber L, Squire IB, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *Eur Heart J*. 2013;34:1404–13. <https://doi.org/10.1093/eurheartj/ehs337>.
 - 46.● Kwon JM, Kim KH, Jeon KH, Park J. Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography. *Echocardiography*. 2019;36:213–8. <https://doi.org/10.1111/echo.14220>. **Interesting analysis exploring in-hospital mortality using only echocardiography results.**
 47. Medved D, Ohlsson M, Höglund P, Andersson B, Nagueh P, Nilsson J. Improving prediction of heart transplantation outcome using deep learning techniques. *Sci Rep*. 2018;8:3613. <https://doi.org/10.1038/s41598-018-21417-7>.
 - 48.● Golas SB, Shibahara T, Agboola S, Otaki H, Sato J, Nakae T, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Med Inform Decis Mak*. 2018;18:44. <https://doi.org/10.1186/s12911-018-0620-z>. **This study presents a unified network model, comparing its performance with other ML methods for the 30-day readmission risk prediction in heart failure patients.**
 - 49.● Chu J, Dong W, Wang J, He K, Huang Z. Treatment effect prediction with adversarial deep learning using electronic health records. *BMC Med Inform Decis Mak*. 2020;20:139. <https://doi.org/10.1186/s12911-020-01151-9>. **In this study, the authors demonstrated the advantage of an adversarial learning scheme to anticipate treatment effects.**
 - 50.● Lewis M, Elad G, Beladev M, Maor G, Radinsky K, Hermann D, et al. Comparison of deep learning with traditional models to predict preventable acute care use and spending among heart failure patients. *Sci Rep*. 2021;11:1164. <https://doi.org/10.1038/s41598-020-80856-3>. **In this study, the authors explored non-traditional outcomes on a large heart failure cohort.**
 - 51.●● Li R, Yin C, Yang S, Qian B, Zhang P. Marrying medical domain knowledge with deep learning on electronic health records: a deep visual analytics approach. *J Med Internet Res*. 2020;22: e20645. <https://doi.org/10.2196/20645>. **In this manuscript, the authors provide an interactive clinical prediction system to visually interpret the prediction results of a deep learning model.**
 - 52.●● Lu XH, Liu A, Fuh SC, Lian Y, Guo L, Yang Y, et al. Recurrent disease progression networks for modelling risk trajectory of heart failure. *PLoS ONE*. 2021;16: e0245177. <https://doi.org/10.1371/journal.pone.0245177>. **In this study, the authors developed an interesting tool for modeling future trajectories of recurrent heart failure.**
 53. Hearn J, Ross HJ, Mueller B, Fan CP, Crowdy E, Duhamel J, et al. Neural networks for prognostication of patients with heart failure. *Circ Heart Fail*. 2018;11: e005193. <https://doi.org/10.1161/CIRCHEARTFAILURE.118.005193>.
 54. Myers J, Arena R, Dewey F, Bensimhon D, Abella J, Hsu L, et al. A cardiopulmonary exercise testing score for predicting outcomes in patients with heart failure. *Am Heart J*. 2008;156:1177–83. <https://doi.org/10.1016/j.ahj.2008.07.010>.
 55. Paulus WJ, Tschope C. A novel paradigm for heart failure with preserved ejection fraction: comorbidities drive myocardial dysfunction and remodeling through coronary microvascular endothelial inflammation. *J Am Coll Cardiol*. 2013;62:263–71. <https://doi.org/10.1016/j.jacc.2013.02.092>.
 56. Pieske B, Tschöpe C, de Boer RA, Fraser AG, Anker SD, Donal E, et al. How to diagnose heart failure with preserved ejection fraction: the HFA-PEFF diagnostic algorithm: a consensus recommendation from the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). *Eur Heart J*. 2019;40:3297–317. <https://doi.org/10.1093/eurheartj/ehz641>.
 57. Segar MW, Patel KV, Ayers C, Basit M, Tang WHW, Willett D, et al. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. *Eur J Heart Fail*. 2020;22:148–58. <https://doi.org/10.1002/ejhf.1621>.
 58. Cohen JB, Schrauben SJ, Zhao L, Basso MD, Cvijic ME, Li Z, et al. Clinical phenogroups in heart failure with preserved ejection fraction: detailed phenotypes, prognosis, and response to spironolactone. *JACC Heart Fail*. 2020;8:172–84. <https://doi.org/10.1016/j.jchf.2019.09.009>.
 59. Pandey A, Kagiyama N, Yanamala N, Segar MW, Cho JS, Tokodi M, et al. Deep-learning models for the echocardiographic assessment of diastolic dysfunction. *JACC Cardiovasc Imaging*. 2021:S1936–878X(21)00355–7. <https://doi.org/10.1016/j.jcmg.2021.04.010>.
 60. Nagueh SF, Smiseth OA, Appleton CP, Byrd BF 3rd, Dokainish H, Edvardsen T, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr*. 2016;29:277–314. <https://doi.org/10.1016/j.echo.2016.01.011>.
 61. Kaptein YE, Karagodin I, Zuo H, Lu Y, Zhang J, Kaptein JS, et al. Identifying Phenogroups in patients with subclinical diastolic dysfunction using unsupervised statistical learning. *BMC Cardiovasc Disord*. 2020;20:367. <https://doi.org/10.1186/s12872-020-01620-z>.
 62. Hedman ÅK, Hage C, Sharma A, Brosnan MJ, Buckbinder L, Gan LM, et al. Identification of novel pheno-groups in heart failure with preserved ejection fraction using machine learning. *Heart*. 2020;106:342–9. <https://doi.org/10.1136/heartjnl-2019-315481>.

63. Chun S, Tu JV, Wijeyesundera HC, Austin PC, Wang X, Levy D, et al. Lifetime analysis of hospitalizations and survival of patients newly admitted with heart failure. *Circ Heart Fail.* 2012;5:414–21. <https://doi.org/10.1161/CIRCHEARTFAILURE.111.964791>.
- 64.● Xiao C, Ma T, Dieng AB, Blei DM, Wang F. Readmission prediction via deep contextual embedding of clinical concepts. *PLoS ONE.* 2018;13: e0195024. <https://doi.org/10.1371/journal.pone.0195024>. **This study exploits Electronic Health Records' data to predict 30-day all-cause readmission of congestive HF patients, identifying 20 subtypes of HF patients.**
- 65.● Awan SE, Bennamoun M, Sohel F, Sanfilippo FM, Dwivedi G. Machine learning-based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics. *ESC Heart Fail.* 2019;6:428–35. <https://doi.org/10.1002/ehf2.12419>. **This study applies deep learning techniques to predict 30-day HF readmission or death in over-65 patients using linked Hospital Morbidity Data Collection.**
- 66.● Allam A, Nagy M, Thoma G, Krauthammer M. Neural networks versus Logistic regression for 30 days all-cause readmission prediction. *Sci Rep.* 2019;9:9277. <https://doi.org/10.1038/s41598-019-45685-z>. **This study aims at predicting 30-day all-cause readmission after HF discharge comparing neural network-based models with logistic regression.**
67. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Winkler S, et al. Telemedical Interventional Management in Heart Failure II (TIM-HF2), a randomised, controlled trial investigating the impact of telemedicine on unplanned cardiovascular hospitalisations and mortality in heart failure patients: study design and description of the intervention. *Eur J Heart Fail.* 2018;20:1485–93. <https://doi.org/10.1002/ejhf.1300>.
68. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JG. Structured telephone support or non-invasive telemonitoring for patients with heart failure. *Cochrane Database Syst Rev* 2015;10:CD007228. <https://doi.org/10.1002/14651858.CD007228.pub3>.
69. Abraham WT, Adamson PB, Bourge RC, Aaron MF, Costanzo MR, Stevenson LW, et al. Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial. *Lancet.* 2011;377:658–66. [https://doi.org/10.1016/S0140-6736\(11\)60101-3](https://doi.org/10.1016/S0140-6736(11)60101-3).
70. Hindricks G, Taborsky M, Glikson M, Heinrich U, Schumacher B, Katz A, et al. Implant-based multiparameter telemonitoring of patients with heart failure (IN-TIME): a randomised controlled trial. *Lancet.* 2014;384:583–90. [https://doi.org/10.1016/S0140-6736\(14\)61176-4](https://doi.org/10.1016/S0140-6736(14)61176-4).
71. Lyngå P, Persson H, Hägg-Martinell A, Hägglund E, Hagerman I, Langius-Eklöf A, et al. Weight monitoring in patients with severe heart failure (WISH). A randomized controlled trial. *Eur J Heart Fail* 2012;14:438–44. <https://doi.org/10.1093/eurjhf/hfs023>.
72. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. *N Engl J Med.* 2010;363:2301–9. <https://doi.org/10.1056/NEJMoa1010029>.
- 73.●● Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan BA, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. *Lancet.* 2018;392:1047–57. [https://doi.org/10.1016/S0140-6736\(18\)31880-4](https://doi.org/10.1016/S0140-6736(18)31880-4). **This study presents the results of a randomised controlled trial aimed at comparing non-invasive multiparameter remote monitoring of HF patients with usual care. The Fontane telemedicine software, which employs business intelligence algorithms, is used to assign patients to a risk category in order to identify patients at higher risk.**
- 74.●● Gontarska K, Wrazen W, Beilharz J, Schmid R, Thamsen L, Polze A. Predicting medical interventions from vital parameters: towards a decision support system for remote patient monitoring. In *International Conference on Artificial Intelligence in Medicine 2021*:293–297. https://doi.org/10.1007/978-3-030-77211-6_33. **This study presents a deep learning method to determine a risk score based on patient vital parameters, which helps practitioners focus their capacities on HF patients at higher risk.**
75. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. *Per Med.* 2018;15:429–48. <https://doi.org/10.2217/pme-2018-0044>.
- 76.●● Stehlik J, Schmalfuss C, Bozkurt B, Nativi-Nicolau J, Wohlfahrt P, Wegerich S, et al. Continuous wearable monitoring analytics predict heart failure hospitalization: the LINK-HF multicenter study. *Circ Heart Fail.* 2020;13: e006513. <https://doi.org/10.1161/CIRCHEARTFAILURE.119.006513>. **This study demonstrates the possibility to monitor exacerbation in HF patients and predicting readmission using a wearable multiparameter sensor.**
77. Henglin M, Stein G, Hushcha PV, Snoek J, Wiltschko AB, Cheng S. Machine learning approaches in cardiovascular imaging. *Circ Cardiovasc Imaging.* 2017;10: e005614. <https://doi.org/10.1161/CIRCIMAGING.117.005614>.
78. Wexler R. When a computer program keeps you in jail: how computers are harming criminal justice. *New York Times* 2017;
79. Varshney KR, Alemzadeh H. On the safety of machine learning: cyber-physical systems, decision sciences, and data products. *Big Data.* 2016;5:246–55.
80. Betancur J, Commandeur F, Motlagh M, Sharir T, Einstein AJ, Bokhari S, al. Deep learning for prediction of obstructive disease from fast myocardial perfusion SPECT: a multicenter study. *JACC Cardiovasc Imaging* 2018;11:1654–1663. <https://doi.org/10.1016/j.jcmg.2018.01.020>.
81. Fraser AG, Byrne RA, Kautzner J, Butchart EG, Szymański P, Leggeri I, et al. Implementing the new European Regulations on medical devices-clinical responsibilities for evidence-based practice: a report from the Regulatory Affairs Committee of the European Society of Cardiology. *Eur Heart J.* 2020;41(27):2589–96. <https://doi.org/10.1093/eurheartj/ehaa382> (PMID: 32484542).
82. Cowan N. The magical mystery four: how is working memory capacity limited, and why? *Curr Dir Psychol Sci.* 2010;19:51–7. <https://doi.org/10.1177/0963721409359277>.
83. Krittanawong C. The rise of artificial intelligence and the uncertain future for physicians. *Eur J Intern Med.* 2018;48:e13–4. <https://doi.org/10.1016/j.ejim.2017.06.017>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.