Future Generation Computer Systems 110 (2020) 802–811

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Deep learning-based cardiovascular image diagnosis: A promising challenge

Kelvin K.L. Wong [a,b,c,*], Giancarlo Fortino [d], Derek Abbott [a,b]

[a] School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide, Australia
[b] Centre for Biomedical Engineering (CBME), The University of Adelaide, Adelaide, Australia
[c] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[d] Department of Informatics, Modeling, Electronics and Systems, University of Calabria, Rende (CS), Italy

A R T I C L E I N F O

Article history:
Received 18 March 2019
Received in revised form 9 August 2019
Accepted 27 September 2019
Available online 3 October 2019

Keywords:
Deep learning
Cardiovascular diagnosis
Image analysis
Artificial intelligence
Big data in medicine

A B S T R A C T

Artificial intelligence (AI) is becoming a vital concept in medicine leading to a rapid emergence of important tools for medical diagnostics. Now, as a crucial machine learning tool in the field of computer vision, deep learning (DL) is being widely used in medical imaging. Furthermore, as reported in the medical literature, DL has been widely used in medical related research. However, the practical application of DL in clinical diagnosis is relatively small, and it is a new field that may have some challenges. How to effectively perform medical image analysis is a major problem in the field of disease diagnosis, and further diagnostic methods need to be developed. At this stage, DL could be viewed as a black box requiring knowledge of its internal workings, and hence presents some crucial technical challenges that need further methodological development. Thereafter with proper diagnostics, pre-operative computerized simulation planning can be carried out for use of appropriate surgical intervention technology. This paper presents important questions on cardiovascular disease (CVD) diagnostics, using this powerful and yet not adequately understood technology. It discusses issues brought by the paradigm shift of AI *vis-à-vis* DL in CVD diagnostics, provides possible solutions to potential issues, and envisions the future of the related machine intelligence applications. The discussed problems are dissected into the modular aspects of DL in relation to CVD image classification, segmentation, and detection. A proper perspective on management of these issues is the key to a successful technological implementation of DL in modern medical science.

1. Introduction

Machine learning [1] is a branch of artificial intelligence (AI) that focuses on how to use experience and improve computing to enhance the performance of computer systems themselves. Machine learning is the process by which computer software is developed to do big data pattern recognition, and be able to continuously learn from and make predictions based on data, and then make adjustments without being specifically programmed to do so. In this way, machine learning effectively automates the process of building analytical models to enable machines to adapt to new scenarios independently. By using knowledge based on experience, machine learning algorithms can be developed to be knowledgeable in making more elaborate and precision diagnosis of disease and predictions of risk of disease. It abandons the artificial input of knowledge to the machine, and instead relies on the algorithm itself to find patterns in the input data. (Refer to Table A.1 of Appendix).

Deep Learning (DL) is a new field in machine learning research. Its motivation is to establish and simulate a neural network for human brain analysis and learning, which mimics the mechanism of the human brain to interpret data [2]. Note that DL takes the original data as the algorithm input, abstracts the original data layer by layer into the final feature representation required by its own task, and ends with the mapping of features to task targets. The entire process is free of any human manipulation. Many research results using DL are analogous to the study of the cognitive principles of the brain, especially the study of visual principles [3,4]. The principle of neural networks is inspired by the physiological structures of the human brain, which uses neurons and synapses as building blocks [5]. Artificial neural networks with deep network structures are the earliest DL models [6]. In a way, DL involves learning about algorithms coded in the human brain to enable learning and knowledge development, for developing more intelligent neural network models. So now, with the rapid development of computer technology and human brain

* Correspondence to: School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide, SA5005, Australia.
E-mail address: kelvin.wong@adelaide.edu.au (K.K.L. Wong).

https://doi.org/10.1016/j.future.2019.09.047
0167-739X/

neuron research, multi-layer perceptron models [7] have been proposed, back-propagation neural network models [8], convolutional neural network (CNN) models [9], deep belief network (DBN) models [10], and other classic models. These research results have greatly promoted the development of DL algorithm architecture, paving the way for its large-scale application in various medical fields (e.g. cardiology [11]).

2. DL-based clinical decision-making research

The effectiveness of the DL model in medicine has led to a wave of data mining and analysis of using this technology in various medical fields, and has also attracted attention in the area of medical image analysis. At present, DL has begun to involve the classification of lesions in medical images [12–14], segmentation [15,16], recognition [17], and brain function studies [18]. Medical images mainly include ultrasound, X-ray, computed tomography (CT), nuclear magnetic resonance (MRI), digital blood vessel silhouette (DSA), and positron emission tomography (PET). In the field of medical image analysis, the main research directions are image segmentation, image registration and information fusion, time series image analysis, and content-based image retrieval. Medical image analysis originally used edge detection, texture features, morphological filtering, as well as the construction of shape models and template matching methods. These analysis methods are usually designed for specific tasks and are known as manual feature design methods.

By contrast DL analyzes tasks in a data-driven manner and automatically learns relevant model features and data characteristics from large datasets of specific problems. The learning process is essentially a solution process for optimization problems. Through learning, the model selects the correct features from the training data, and thereafter enabling the classifier to make the right decisions when testing new data [19,20]. In recent years, DL algorithms have made some breakthroughs in the field of medical image analysis. One good example is the successful implementation of machine learning by the Google AI team to perform medical diagnosis, where their algorithms perform well equivalent to medical experts [21].

In fact, Google has been building one of the largest neural networks yet, with more than a billion connections, resulting in huge improvement in image recognition. However, can deep learning move artificial intelligence toward something rivaling human intelligence? In this regard, our task is to know how the brain works, and how it might provide a guide to building intelligent machines, Now to build real intelligent machines, we need to also build into the neural networks the concept of time, by which the artificial brain can mimic the human brain in better recording sequences of patterns to thereby properly analyzing motion of objects (such as heart motion), in order to make precision diagnosis of the heart. For example, a myocardial ischemic left ventricle undergoes remodeling and less motion from diastole to systole compared to that of the normal ventricle. So then our artificial brain (or super intelligent machine) needs to be designed to also analyze sequences of patterns, towards making more accurate diagnosis.

3. The role of DL in CVD diagnosis

Cardiovascular disease (CVD) is a common disease that endangers human health, involving high blood pressure, coronary heart disease, rheumatic heart disease and cerebrovascular disease. Imaging diagnosis of CVD mainly relies on detection methods *vis-à-vis* cardiac ultrasound, CV angiography, and CV magnetic resonance and computed tomography [22]. Moreover, CV medical imaging has become an integral part of the diagnosis and treatment of CVD, and is increasingly important. The common types of tasks currently applied in the field of CV medical image analysis are classification, detection and segmentation. With the rapid development of medical imaging technology, CV medical image analysis has entered the era of big data, as to how to extract useful knowledge from a large number of CV medical images and provide a more accurate basis for clinical diagnosis.

However, traditional pattern recognition or machine learning methods applied to CV image analysis require *a priori* extraction from the original data to train the learning model. Due to difficulty in feature selection, the model may have over-fitting problems, and the generalization ability is difficult to guarantee. In the last two decades, the DL methods in the field of computer vision have become increasingly mature, and DL algorithms have fully demonstrated the potential to solve the dilemma faced by traditional CV image analysis methods. The application of DL now provides a new opportunity for automated analysis of CV images, and for assisting medical experts in achieving high-precision intelligent diagnosis of CVD. A summary of the DL frameworks applied in CVD can be referred to in Table A.2 of Appendix).

4. Deep learning-based methods

4.1. Classification

Image classification is one of the first application of DL [11], which has made significant contributions to medical image analysis. In a classification task, one or more images are typically used as inputs, and diagnostic variables (such as the presence or absence of a disease) constitute the output. By comparing logistic regression [23], the performance of classification and regression trees and neural network classification techniques can predict the presence of coronary artery disease (CAD) [24]. Experiments have shown that multi-layer perceptron (MLP) analysis is the best technique for predicting CAD from a coronary angiography dataset in lieu of human observation. A striking feature of neural networks is their ability to reproduce the nonlinear relationship between possible signs and symptoms and the diagnosis of CAD. Note that an MLP is a feed-forward neural network trained using backpropagation algorithms. It is a supervised network, with one or two hidden layers, which can approximate almost any input–output mapping. In addition, MLP uses the gradient descent learning theory, and a back-propagation algorithm can calculate the sensitivity of the cost function with respect to each weight in the network. This neural network topology can be used to obtain high accuracy in diagnosing coronary heart disease (CHD) [25]. An advancement of the neural framework known as the Convolutional Neural Network (CNN) is built upon the MLP architecture by exploitation of strong spatially local correlation present in the medical images (Fig. 1).

To assist clinicians in the diagnosis heart disease, the CNN can be applied to the classification of echocardiographic video images [26]. Compared with MLP, CNN has closer resemblance to an actual biological neural network, and its special structure of local weight sharing has unique advantages in image processing. In addition, the sparse connection between layers in CNNs greatly reduces the number of parameters, thereby effectively shortening the network convergence time. The fusion neural network (FNN) architecture combines the spatial and temporal information of echocardiography, designing two CNN networks along the spatial and temporal directions and performing them separately, while the integration of spatial and temporal information is achieved by class scores obtained from the two networks [27]. Unlike the MLP model, the FNN model is not only superior to traditional manual methods, but also superior to other existing techniques that process the non-standardized dataset. The DL algorithm can

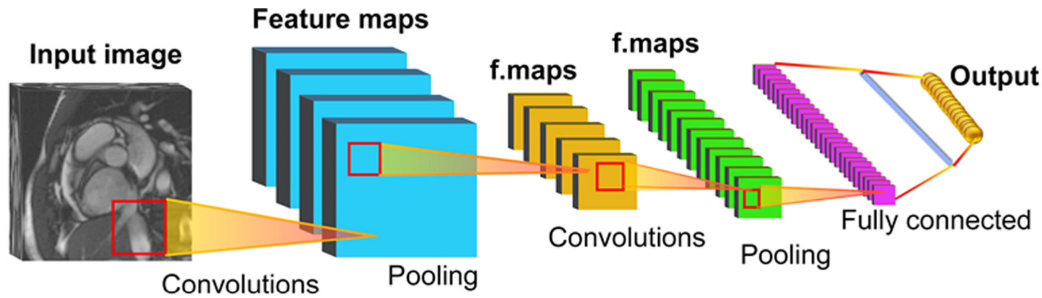


Fig. 1. Typical CNN architecture comprising a set of learnable filters (or kernels), which have a small receptive field, but extend through the full depth of the input volume so as to allow superior generalization on the medical image vision issues.

classify any electrocardiographic video without the availability of ECG data, which provides significant benefits in the development of computer-aided diagnostic systems.

Although the FNN model has achieved useful classification performance, it is still necessary to manually extract the features of the image, and thus this method has room for further improvement. Therefore, a new DL method that does not require manual feature extraction has been developed to classify plaque types in clinical intravascular optical coherence tomography (IVOCT) images of coronary arteries [28]. By introducing a fully connected conditional random field (CRF) method to normalize the classification of large areas and using cross-validation procedures to evaluate the performance of the classifier, this system can achieve a higher classification accuracy in an exponentially large number of images (i.e. 10^n, for $n > 2$). Inspired by a deep fine-tuning network, a combined fine-tuning model for automatic classification of the coronary arteries began to emerge [29]. Therein, the imaging techniques used in the datasets employed for model training and testing are the same as in a deep fine-tuning network, and have been beneficial for intracoronary tissue imaging pertaining to pediatric patients that have been tested clinically in recent years. By fine-tuning the pre-training network from the classification layer to the third convolutional layer, the CNN is used as a feature extractor for each image sequence of each patient, and then used before the classification layer. Finally, the random forest performs classification of the coronary layer by applying the extracted features.

4.2. Segmentation

At present, the application of DL in cardiac image segmentation mainly focuses on left/right ventricular (LV/RV) segmentation and CV segmentation. Due to the traditional active contours [30], the deformable templates technique [31] is only effective if it has prior knowledge of the shape and appearance of the heart ventricle. A combined DL model that addresses this problem has therefore emerged, which combines neural networks with improved search techniques for robust left ventricular segmentation from ultrasound data [32]. By applying multi-layer Restricted Boltzmann Machines (RBM), we can implement a complex, fully connected Boltzmann machine that is known as Deep Boltzmann Machine [33], which is shown by Fig. 2. Using this technique, we can achieve robust segmentation in spite of poor image quality in training data. This neural network is characterized by its ability to be used for dimensionality reduction and learning. Since the search in the parameter space is complex and time consuming, the search process can be simplified by reducing the search space and running the search process on image pyramid using a classifier for each image scale.

Based on the successful application of RBM, the Deep Belief Network (DBN) was introduced for segmentation of the left ventricle from ultrasound data, and it was constructed by multiple RBMs [34]. In the segmentation process, DBN was first used to pre-train unlabeled data layer by layer, and then train a limited number of labeled 2D ultrasound images. This method effectively solves the problem of insufficient datasets with annotations. The approach also solves the problem of rigid and non-rigid detection by combining DL methods that simulate left ventricular (LV) appearance with derivative-based search algorithms. In addition, the DBN is also used to detect the appearance and boundaries of the LV and then the Active Shape Model (ASM) is used for segmentation. A combination of the active contour model (distance regularization level set) with DBN was proposed for the study of LV segmentation in magnetic resonance (MR) images [35]. In subsequent work, this research was extended to the segmentation of the right ventricle (RV) of the heart [36].

Combination of DL algorithms and deformable models can be used to develop and evaluate fully automated LV segmentation tools based on short-axis cardiac MR image datasets [37]. We note that a convolutional network (CNN) based method was used to detect the LV chamber in the MRI dataset, and then exploited a stacked autoencoder to generate the initial shape of the LV. Finally, the results generated by the encoder are combined into the deformable model to improve the accuracy and robustness of the segmentation. Unlike the probability maps of the DBN model output processed by the level set in the traditional multi-level CNN method [34,35], the level set was directly used for the initial image. The performance of the fully automatic segmentation method of the left ventricle in cardiac MRI is basically consistent with ground truth, and the verification metric reaches 96.69%. Subsequently, a LV auto-segmentation method based on simplified pulse coupled neural network (SPCNN) and an a priori constrained GVF [38] emerged. Next, a time domain method was first designed to extract the rectangular region around the heart, and then use the SPCNN to detect the LV. Furthermore, SPCNN can implement self-correction segmentation. The pericardium is then accurately segmented based on maximum gradient search and gradient vector flow (GVF).

In contrast to the non-end-to-end combined depth model [35, 37], the fully convolutional neural networks (FCNs) can segment the LV and RV, as well as the myocardium more efficiently [39]. The traditional CNN model is only applicable to one or two steps of the ventricular segmentation task, whereas the FCN model can be trained end-to-end without any pre-processing or post-processing of the image. The FCN classifies the image at the pixel level. Unlike the classical CNN, which uses a fully connected layer to obtain fixed-length feature vectors after classification, the FCN can accept input images of any size. By using the deconvolution layer to upsample the feature map of the last convolutional layer to restore it to the same size of the input image, a prediction can be generated for each pixel while preserving the space information in the original input image, and finally for pixel-by-pixel classification on the upsampled feature map (Fig. 3). FCN has been used to segment the LV on 2D ultrasound images [39].

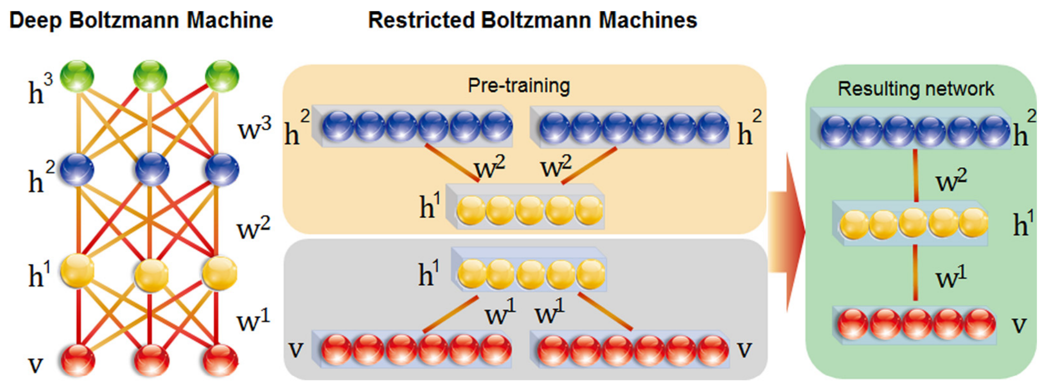


Fig. 2. Multilayer Deep Boltzmann Machine implements and learns a stack of modified RBM's for pre-training, wherein $\{v, h^1, h^2\}$ is defined as the energy state and $\{W^1, W^2\}$ represents visible-to-hidden and hidden-to-hidden symmetric interaction terms.

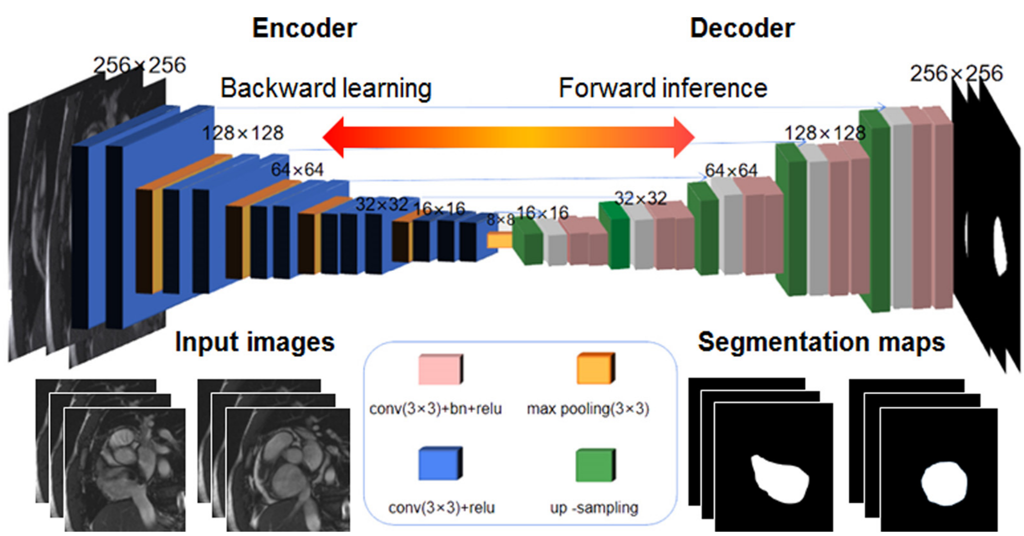


Fig. 3. Segmentation network based on Fully Convolutional Networks is built upon locally connected layers, such as convolution, pooling and up-sampling. The network uses the down-sampling path by capturing contextual information from the raw images (input) and the up-sampling path by recovering spatial information to obtain the segmentation maps (output).

The iterative multi-domain regularized DL method effectively solves the problem of insufficient training data. In subsequent studies, researchers began to turn their attention to 3D images that contained more information than 2D images. Due to the high complexity of parameterization in 3D images, the edge space deep learning (ESDL) method has emerged. By taking advantage of efficient object parameterization in hierarchical marginal space and the advantages of automated feature design of DL network architecture, this method achieves the efficiency and robustness of segmentation in 3D medical images.

4.3. Detection

As one of the important prerequisites for a treatment plan, the purpose of the detection task is to find the area associated with the symptoms of the disease. In response to the low resolution of low-dose chest CT, a new method for the automatic detection of low-dose chest CT coronary, thoracic aorta and valvular calcification was established by using two consecutive convolutional neural networks (CCNN) [40]. The first CNN identifies and marks potential calcification based on its anatomical location, specifically, using a stacked expanded convolution to promote large receptive fields, which enables spatial markers of high-density voxels to be identified and implemented. The second CNN then identifies the true calcification in the detected candidate. Only voxels classified as calcium by the first CNN are classified as true-positive or false-positive by the second CNN. In other methods of performing calcification detection tasks, sensitive areas in the image are typically detected first. The CCNN does not need to manually extract and input explicit spatial features, and the spatial background in the three orthogonal 2D patches is able to be successfully recognized by CNN with expanded convolution characteristics [40].

Although the above mentioned CCNN has achieved working results, it does not make full use of the spatiotemporal information of image data, and there is room for further improvement. In contrast to CCNN, a combination of convolutional neural networks and circulating neural networks was proposed to detect and characterize coronary plaque types [41]. This method also considered temporal and spatial information contained in multi-planar reformatted (MPR) images of coronary arteries, and did not require manual feature extraction. Features can be extracted from the coronary arteries by using a 3D convolutional neural network, and then aggregating these extracted features by performing recursive neural networks of two simultaneous multi-class classification tasks. Differently from most traditional methods that rely on coronary luminal segmentation to detect and characterize coronary plaque and stenosis, this method only requires extraction of the coronary artery centerline from coronary CT angioplasty (CCTA) as input. The method automatically classifies patients into (i) patients without coronary plaque, and (ii) patients with coronary plaques and stenosis who require further CV examination.

In clinical practice, fractional flow reserve (FFR) measurements are often used during invasive coronary angiography (ICA) to determine the functional significance of coronary stenosis. Therefore, methods such as detecting coronary plaque and stenosis similar to ESDL method cannot effectively reduce the number of ICA procedures, whereas CNN can be used to detect LV myocardium in CCTA to automatically identify patients with significant coronary stenosis [42]. This CNN model can detect and segment the myocardial portion of the LV, to be then encoded using an unsupervised convolution autoencoder (CAE) to characterize the segmented LV myocardium. Next, the features extracted by support-vector machine (SVM) classification are used according to the existence of a functionally significant stenosis. Automated detection of the LV myocardium in the CCTA scan can be used to identify patients with functionally significant coronary stenosis without assessing coronary anatomy. In the field of CV image detection, researchers have now even extended their DL methods to the detection of vascular defects [43], cardiac three-dimensional ultrasound catheter testing [44] and other fields.

5. Perspectives and identified problems

5.1. Classification

Combining the recent research results based on neural networks for CV image classification [24,27], it is clear that the accuracy of DL exceeds that of the traditional machine learning and pattern recognition methods and is continuously being improved. The balance and scale of the dataset largely influence the generalization of the neural network model. All of the above-mentioned studies are able to mitigate the impact of a smaller training set by using a combined or regularized approach. More strikingly, these studies show that DL algorithms that do not require manual extraction, and the input of explicit spatial features are more focused on the application of advanced classifications to directly perform lesion characterization without pre-processing steps that use lesion segmentation.

Although DL has achieved remarkable results in the field of CV medical image classification, will it achieve the same effect in clinical diagnostics? The DL model applied to medical image analysis can have research and clinical practice value only when it simultaneously satisfies high robustness, accuracy, and timeliness. Radiologists initially refer to non-imaging information, such as clinical history, to help them diagnose the disease. These sources of information are especially important in distinguishing between similar diseases. The current application of DL to diagnose CVD mainly relies on a single set of medical imaging data, lacking genetic factors and other relevant clinical parameters. If a human medical expert with extensive diagnostic experience cannot make elaborate diagnosis based on a single image data, then how could an artificial intelligence system do so? The DL algorithms may uncover subtleties in medical images that are overlooked by human physicians, and it is a challenge on how they can have a significant impact on diagnostics. Thus, the actual effect of DL on clinical diagnostic classification is still in the making, and somewhat dilemmatic. In addition, the credibility and scale of the dataset is also an issue.

One can figure out the real root cause: the deviation of accuracy caused by information asymmetry and inadequacy during the process of data acquisition and data-basing [45]. Therefore, DL requires massive amounts of data to *train* the classifier on the *training* set, and test it on a hold-out test, to achieve useful performance. The labeling of medical image data is performed by an authoritative medical expert group based on the patient's radiology report and other relevant clinical parameters, which deters collection of large-scale medical image data. Due to the complexity of CV-related diseases, it may be difficult to determine the defined class based on visual information, for example in discerning the normal heart and the enlarged heart in an X-ray image. Therefore, these labeling operations may require more AI intensive analyses of the images, to promote the accuracy of the tags. However, with the further improvement of the performance of the testing equipment, these problems could be solved. One can then obtain larger and more accurate medical image datasets, and image processing methodologies to train the DL network to obtain more authoritative classification-based diagnostics results.

5.2. Segmentation

Due to the rich information content in medical images, it is cumbersome and time consuming for radiologists to identify anatomical structures and perform measurements so as to obtain quantification for the diagnostics. The above-mentioned studies have shown that the use of automated or semi-automated CV medical image analysis systems has some unparalleled advantages in reducing inspection time, increasing the reliability of examinations, and improving the diagnostic accuracy of physicians. Recent advances in DL have shown that compared to the use of non-learning data-driven methods *vis-à-vis* the methods for simple thresholding and region growing, deep neural networks can successfully classify patient data and obtain highly accurate segmentation results.

The emergence of DL effectively solves the problems faced by non-learning-based data-driven methods and improves the accuracy of segmentation. Nevertheless, will applications of this type of technique in the field of CV image segmentation present other major challenges for computer vision researchers? The latest results of DL mainly focus on 2D image segmentation, and there is less attention given to 3D image segmentation. This shows that 3D DL is still a challenging task. Why is this happening? Firstly, a zero-sum attempt at containment of human intervention is likely to lead to increase in computational expense during the implementation of classification machines in medical diagnostics. For real clinical applications, the performance, the timeliness, and the computational costs are the key issues to be considered. However, evaluating deep neural networks applied to large 3D images does not easily meet the requirements. Secondly, since 3D images contain more information, so networks that use 3D patches as input require more training data.

Due to the difficulty of manually labeling 3D image data and the sharing of patient image data, medical imaging researchers often explore limited training samples with only a few thousands of images. However, the transition to 3D image analysis is an inevitable trend in the field of CV image segmentation. In addition, in clinical practice, doctors need to estimate cardiac parameters such as cardiac ejection fraction, volume and mass by segmentation methods. Since different models of MRI scanners have different acquisition protocols, the acquired cardiac MRI images will vary greatly. In the context of DL methods, this poses a challenge for the clinical application and the promotion of DL algorithms. At present, we need to improve the theoretical performance of the algorithms, and also take into account the complex factors in clinical examination.

5.3. Detection

The field of medical image analysis has begun applying DL technology to solve key issues in image detection. However, the transition from an artificial-based system to a system that learns

Table A.1
Summary of relationships for machine learning techniques (along with their brief descriptions and modeling characteristics).

Technique	Brief description	Biologically inspired?	Modeling characteristics	Learning characteristics	Core formulas
Neural networks	Network or circuit of artificial neurons (or nodes) and synapses.	Yes	Error-correction learning based modeling	Supervised learning	$f(x) = o = w_0 + \sum_{i=1}^{n} w_i x_i$
Decision trees	A tree-like model of decisions and their possible consequences.	No	Decision rule modeling	Supervised learning	$Entropy = \sum_{v=0}^{1} -P \cdot \log(P)$ $InfoGain = P_+ \cdot [-P_{+t} \cdot \log(P_{+t}) - P_{+(t-1)} - \cdot \log(P_{+(t-1)})]$
Random forests	A multitude of decision trees at training time and outputting the classification or regression of the individual trees	No	Decision rule modeling	Supervised learning	$a_n \leq N_n(x, \Theta) \leq b_n \text{ and } a_n \leq \frac{1}{M} \sum_{m=1}^{M} N_n x, \Theta_m \leq b_n$ $\lvert m_{M,n}(x) - \tilde{m}_{M,n}(x)\rvert \leq \frac{b_n - a_n}{a_n} \tilde{m}_{M,n}(x)$
Associations and sequence discovery	Rule-based method for discovering relations between different variables.	No	Decision rule modeling	Supervised learning	$\sup\ p(X) = \frac{\lvert\{t \in T; X \subseteq t\}\rvert}{\lvert T\rvert}$ $conf(X \Rightarrow Y) = \sup\ p(X \cup Y)/\sup\ p(X)$ $lift(X \Rightarrow Y) = \frac{\sup\ p(X \cup Y)}{\sup\ p(X) \times \sup\ p(Y)}$
Gradient boosting and bagging	Model is constructed in a stage-wise fashion, and is generalized by allowing optimization of an arbitrary differentiable loss function.	No	Decision rule modeling	Ensemble learning	$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)$ $F_m(x) = F_{m-1}(x) = \arg\min_{h_m \in \mathrm{H}} \left[\sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + h_m(x_i)) \right]$
Support vector machines	Supervised learning models with associated learning algorithms that perform classification and regression analysis.	No	Regression-based modeling	Supervised learning	$f(x) = sign[\lambda \cdot y \cdot K(x_i \cdot x_j)]$ $K(x_i \cdot x_j) = \exp \left[-\sqrt{\frac{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{width_{hist}}} \right]$
Nearest-neighbor mapping	A type of proximity map that analyzes relationships between two datasets based on 'nearest neighbors' concept.	No	Proximity-based decision modeling	Unsupervised learning	$\hat{f}(x) \leftarrow \frac{\sum f(x)}{k}$ $DE(x_i, x_j) = \sqrt{(x_i - x_j)^2 + (y_{xi} - y_{xj})^2}$
k-means clustering	Aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.	No	Proximity-based decision modeling	Unsupervised learning	$\arg\ \min \sum_{i=1}^{k} \sum_{x \in S_i} \lVert x - \mu_i \rVert^2 = \arg\ \min \sum_{i=1}^{k} \lvert S_i \rvert VarS_i$ $\arg\ \min \sum_{i=1}^{k} \frac{1}{2\lvert S_i\rvert} \sum_{x,y \in S_i} \lVert x - y \rVert^2$
Self-organizing maps	A low-dimensional, discretized representation of the input space of the training samples, called a map, is constructed.	No	Competitive learning based modeling	Unsupervised learning	$Wv(s+1) = Wv(s) + \theta(u, v, s) \cdot \alpha(s) \cdot (D(t) - Wv(s))$
Local search optimization techniques (e.g., genetic algorithms)	A heuristic algorithm that finds the global minimum in order to solve nonlinear or non-differentiable optimization problems.	Yes	Evolutionary programming.	Unsupervised learning	$\begin{cases} \max\ f(x) \\ X \in R \\ R \subset U \end{cases}$
Expectation maximization	Finds the maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables.	No	Statistical-based modeling	Unsupervised learning	$\hat{\sigma}_w^2 = \frac{1}{N} \sum_{k=1}^{N} (\hat{x}_{k+1} - \hat{F}\hat{x}_k)^2$ $\hat{F} = \frac{\sum_{k=1}^{N} (\hat{x}_{k+1} - \hat{F}\hat{x}_k)}{\sum_{k=1}^{N} \hat{x}_x{}^2}$

(continued on next page)

Table A.1 (*continued*).

Technique	Brief description	Biologically inspired?	Modeling characteristics	Learning characteristics	Core formulas
Multivariate adaptive regression splines	A non-parametric regression technique based on an extension of linear models that models nonlinearities and interactions between different variables.	No	Regression-based modeling	Unsupervised learning	$\hat{f}(x) = \sum_{i=1}^{k} c_i B_i(x)$
Bayesian networks	Implements Bayesian inference for probability computations to model conditional dependence.	No	Probabilistic modeling	Supervised learning	$tuples\neg\ for\ y = 0 \wedge y = 1$
Kernel density estimation	A fundamental data smoothing framework that estimates the probability density functions of a random variable.	No	Probabilistic modeling	Unsupervised learning	$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h})$
Principal component analysis	Implements an orthogonal transformation to convert a set of observations of possible correlated variables into a set of linearly uncorrelated variables.	No	Statistical-based modeling	Unsupervised learning	$Eigenvalue = [A] - \lambda I$ $Eigenvector = Eigenvalue \cdot [A]$ $f(x) = Eigenvector^T \cdot [x_{i1} \ldots x_{jn}]$
Singular value decomposition	Generalization of the Eigen-decomposition of a positive semidefinite normal matrix to any matrix via an extension of the polar decomposition.	No	Linear transformation	Unsupervised learning	$M = U\Sigma V^*$ $M_v^\rho = \sigma_u^\rho\ and\ M^{*\rho}_u = \sigma_v^\rho$
Gaussian mixture models	Assumes all the data points are generated from a consolidation of Gaussian distributions with unknown parameters.	No	Probabilistic modeling	Unsupervised learning	$P(x\|\bar{x}) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{1}{2}(\frac{x - \bar{x}}{\sigma})^2\right]$ $Z_{cs} = \frac{N_A C_B + N_B C_A}{N_A + N_B}$ $P(Z_{cs}) \to 0.50$
Sequential covering	Repeatedly learns a single rule to create a decision list (or set) that encompasses the entire dataset rule by rule.	No	Decision rule modeling	Unsupervised learning	$R = 2\sum_{i=1}^{k} f_i \log(f_i/e_i)$ $POIL = p_1 \times (\log_2 \frac{p_1}{p_1 + n_1} - \log_2 \frac{p_0}{p_0 + n_0})$

features from data is a gradual process. Accurate positioning of specific biomarkers or anatomical structures in medical images is of great importance in clinical treatment, which is directly related to the quality of the treatment. Recent research has shown that the use of end-to-end automated CV image detection algorithms based on DL has outstanding advantages in identifying quantitative lesion areas and improving the diagnostic efficiency of diseases [40–44]. As research deepens, DL methods are being applied to tools for detecting and locating surgical videos [46,47] and have been successful in clinical practice [48]. In future studies, DL methods can even be applied to cardiac interventions to reduce the surgeon's proficiency requirements, while meeting the acceptable criteria for performing the procedure, and to expand implementation of cardiac interventions in underdeveloped areas of medical resources.

In recent years, with the deep improvement of image classification accuracy, the medical image detection algorithm based on DL has gradually become the mainstream. However, is the current DL-based detection algorithm capable of processing all types of image data in the CV field? Firstly, the DL framework has difficulty finding the discriminating information that exists in the image patch without local annotation. Secondly, there is a lack of sufficient research for detecting sparsely distributed objects (such as arterioles and capillaries) from large-scale medical images using DL methods. These challenges all affect the performance of the framework in terms of speed and accuracy of detection. These statements sounded ominously like an early bugle-call in the slow replacement of humans even for the labeling of data due to possible erroneous diagnosis, which can only be refined overtime by AI resulting in full replacement of humans by robots as the gold standard in future. As the performance of hardware devices increases, the amount of computation within a certain range will no longer become a constraint for real-time detection.

6. Existing challenges

The field of medical image analysis has begun to apply DL technology to address key CV imaging issues. However, the transition from an artificial-based system to a system that learns features from data is a gradual process. It is notable that DL techniques appear to have achieved considerable performance in many medical application studies. Nevertheless, AI specialists have a lot of challenges in using DL as a new tool for diagnosing vital CVD. It is noteworthy that DL technology brings new methods to the most important feature extraction of machine learning. The current application of this advanced technology in clinical practice still faces some serious challenges:

Table A.2
Overview of DL frameworks applied in cardiovascular imaging studies (along with their abbreviations)

	Purpose	Imaging technique	ML technique	Abbreviation	Studies performed
Classification	Predicting coronary artery disease	Coronary angiography	Multi-layer perceptron	MLP	Kurt et al. [24] Atkov et al. [25]
	Classification of echocardiographic video images	Echocardiography	Convolutional neural networks	CNN	Madani et al. [26]
	Classification of echocardiographic video images	Echocardiography	Fusion neural network	FNN	Gao et al. [27]
	Plaque classification	Intravascular optical coherence tomography	Fully connected conditional random field	CRF	Kolluru et al. [28] Abdolmanafi et al. [29]
Segmentation	Segmentation of the left ventricle (LV)	Echocardiography	Deep belief network	DBN	Carneiro et al. [32] Carneiro et al. [34]
	Segmentation of the right ventricle (RV)	Magnetic resonance imaging	Deep belief network	DBN	Ngo et al. [35] Zotti et al. [36]
	Segmentation of LV, RV and myocardium	Echocardiography	Fully convolutional neural networks	FCN	Jang et al. [39]
	Segmentation of LV	Magnetic resonance imaging	Simplified pulse coupled neural network	SPCNN	Ma et al. [38]
Detection	Detection of coronary, thoracic aorta and valvular calcification	Computed tomography	Consecutive convolutional neural networks	CCNN	Lessmann et al. [40]
	Detection of LV myocardium	Coronary computed tomography angioplasty	Convolutional neural networks	CNN	Zreik et al. [41,42]

1. There is a lack of high-quality labeled training samples for model training, so the trained models may be over-fitting or poorly generalized. For this reason, it is necessary to raise the generalization of the model in various situations, which is cumbersome and time consuming.
2. The model obtained by a DL method can be likened to a black box, which can be viewed in terms of its inputs and outputs, without knowledge of its internal workings. Currently, no DL network can adequately explain its decision-making process, so the acceptance in the medical industry is also a problem.
3. There are legal and ethical issues with the use of clinical images in commercial systems, and failure to use these reliable data can hinder the performance of DL models.

It is important to recognize this tension between complexity and simplicity: Deep learning is only a representative of the connected school in several schools of artificial intelligence. It is not necessarily the best algorithm, and its performance limit needs to be reasonably evaluated. There is usually a robust solution to most human imaging problems, for which artificial intelligence holds a promising challenge. Whether it is neat and plausible depends on the data it relies upon for learning. In addition, the current research on DL in the field of medical image analysis use evaluation indicators from the computer field, which needs adequate medical understanding in order to be incorporated into clinical practice.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (grant number 81771927).

Appendix

See Tables A.1 and A.2.

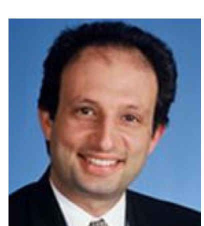
References

[1] P. Langley, The changing science of machine learning, Mach. Learn. 82 (2011) 275–279.
[2] Y. Bengio, Y. LeCun, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[3] D.H. Hubel, T.N. Wiesel, Receptive fields of cells in striate cortex of very young, visually inexperienced kittens, J. Neurophysiol. 26 (1963) 994–1002.
[4] S. Le Vay, T.N. Wiesel, D.H. Hubel, The development of ocular dominance columns in normal and visually deprived monkeys, J. Comp. Neurol. 191 (1980) 1–51.
[5] K.O. Stanley, J. Clune, J. Lehman, R. Miikkulainen, Designing neural networks through neuroevolution, Nat. Mach. Intell. 1 (2019) 24–35, http://dx.doi.org/10.1038/s42256-018-0006-z.
[6] J.E. Dayhoff, J.M. DeLeo, Artificial neural networks: opening the black box, Cancer 91 (2001) 1615–1635.
[7] Y. Chen, F. Bastani, in: IEEE International Conference on Systems, Man and Cybernetics, 1990, 370-372.
[8] Y. LeCun, et al., Adv. Neural Inf. Process. Syst. (1990) 396–404.
[9] P. Simard, L. Bottou, P. Haffner, Y. LeCun, Adv. Neural Inf. Process. Syst. (1999) 571–577.
[10] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (2006) 1527–1554.
[11] S.J. Al'Aref, et al., Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. http://dx.doi.org/10.1093/eurheartj/ehy404, 2018.
[12] S.M. Plis, et al., Deep learning for neuroimaging: a validation study, Front. Neurosci. 8 (2014) 229.
[13] N. Tajbakhsh, et al., Convolutional neural networks for medical image analysis: Full training or fine tuning?, IEEE Trans. Med. Imaging 35 (2016) 1299–1312.
[14] S. Parisot, et al., Disease prediction using graph convolutional networks: Application to autism spectrum disorder and alzheimer's disease, Med. Image Anal. (2018).
[15] S.C. Turaga, et al., Convolutional networks can learn to generate affinity graphs for image segmentation, Neural Comput. 22 (2010) 511–538.
[16] H.R. Roth, et al., in: International Conference on Medical Image Computing and Computer-assisted Intervention, 2015, 556-564.
[17] H.R. Roth, et al., in: International Conference on Medical Image Computing and Computer-assisted Intervention, 2014, 520-527.
[18] I. Kuzovkin, et al., Activations of deep convolutional neural network are aligned with gamma band activity of human visual cortex, Commun. Biol. 1 (2018) 107, http://dx.doi.org/10.1038/s42003-018-0110-y.
[19] H. Zenil, N.A. Kiani, A.A. Zea, J. Tegnér, Causal deconvolution by algorithmic generative models, Nat. Mach. Intell. 1 (2019) 58–66, http://dx.doi.org/10.1038/s42256-018-0005-0.
[20] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436, http://dx.doi.org/10.1038/nature14539.
[21] V. Gulshan, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, JAMA 316 (2016) 2402–2410.

[22] J.P. Greenwood, et al., Cardiovascular magnetic resonance and single-photon emission computed tomography for diagnosis of coronary heart disease (CE-MARC): a prospective trial, Lancet 379 (2012) 453–460.
[23] D.W. Hosmer, S. Lemeshow, Applied Logistic Regression, Wiley, 2000.
[24] I. Kurt, M. Ture, A.T. Kurum, Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease, Expert Syst. Appl. 34 (2008) 366–374.
[25] O.Y. Atkov, et al., Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters, J. Cardiol. 59 (2012) 190–194.
[26] A. Madani, R. Arnaout, M. Mofrad, R. Arnaout, Fast and accurate view classification of echocardiograms using deep learning, Digit. Med. 1 (2018) 6.
[27] X. Gao, W. Li, M. Loomes, L. Wang, A fused deep learning architecture for viewpoint classification of echocardiography, Inf. Fusion 36 (2017) 103–113.
[28] C. Kolluru, et al., Deep neural networks for A-line-based plaque classification in coronary intravascular optical coherence tomography images, J. Med. Imaging 5 (2018) 044504.
[29] A. Abdolmanafi, L. Duong, N. Dahdah, F. Cheriet, Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography, Biomed. Opt. Express 8 (2017) 1203–1220.
[30] O. Bernard, B. Touil, A. Gelas, R. Prost, D. Friboulet, in: IEEE International Conference on Image Processing (ICIP), 2007, III-157-III-160.
[31] T. Chen, J. Babb, P. Kellman, L. Axel, D. Kim, Semiautomated segmentation of myocardial contours for fast strain analysis in cine displacement-encoded MRI, IEEE Trans. Med. Imaging 27 (2008) 1084–1094.
[32] G. Carneiro, J. Nascimento, A. Freitas, in: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2010, 1085-1088.
[33] R. Salakhutdinov, G. Hinton, in: International Conference on Artificial Intelligence and Statistics, 2009, 448-455.
[34] G. Carneiro, J.C. Nascimento, A. Freitas, The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods, IEEE Trans. Image Process. 21 (2012) 968–982.
[35] T.A. Ngo, Z. Lu, G. Carneiro, Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance, Med. Image Anal. 35 (2017) 159–171.
[36] C. Zotti, Z. Luo, A. Lalande, O. Humbert, P.-M. Jodoin, Novel Deep Convolution Neural Network Applied to MRI Cardiac Segmentation. arXiv preprint arXiv:1705.08943, 2017.
[37] M. Avendi, A. Kheradvar, H. Jafarkhani, A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI, Med. Image Anal. 30 (2016) 108–119.
[38] Y. Ma, et al., An SPCNN-GVF-based approach for the automatic segmentation of left ventricle in cardiac cine MR images, Int. J. Comput. Assist. Radio. Surg. 11 (2016) 1951–1964.
[39] Y. Jang, Y. Hong, S. Ha, S. Kim, H.-J. Chang, in: International Workshop on Statistical Atlases and Computational Models of the Heart, 2017, 161-169.
[40] N. Lessmann, et al., Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions, IEEE Trans. Med. Imaging 37 (2018) 615–625.
[41] M. Zreik, et al., A recurrent CNN for automatic detection and classification of coronary artery plaque and stenosis in coronary CT angiography, IEEE Trans. Med. Imaging (2018) http://dx.doi.org/10.1109/TMI.2018.2883807.
[42] M. Zreik, et al., Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis, Med. Image Anal. 44 (2018) 72–85.
[43] S. Filippov, A. Moiseev, A. Andrey, Deep Diagnostics: Applying Convolutional Neural Networks for Vessels Defects Detection. arXiv preprint arXiv:1705.06264, 2017.
[44] H. Yang, C. Shan, A.F. Kolen, P.H. de With, in: IEEE International Conference on Image Processing (ICIP), 2018, 371-375.
[45] E. Begoli, T. Bhattacharya, D. Kusnezov, The need for uncertainty quantification in machine-assisted medical decision making, Nat. Mach. Intell. 1 (2019) 20–23, http://dx.doi.org/10.1038/s42256-018-0004-1.
[46] R. Girshick, in: IEEE International Conference on Computer Vision, 2015, 1440-1448.
[47] D. Sarikaya, J.J. Corso, K.A. Guru, Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection, IEEE Trans. Med. Imaging 36 (2017) 1542–1549.
[48] A.P. Twinanda, et al., Endonet: A deep architecture for recognition tasks on laparoscopic videos, IEEE Trans. Med. Imaging 36 (2017) 86–97.

Kelvin Wong received the B.Eng. degree (Hons.) in mechanical and production engineering from Nanyang Technological University in 2001, the M.A.I.T. degree in applied information technology from The University of Sydney in 2003, and the Ph.D. degree in electrical and electronic engineering from The University of Adelaide in 2009. From 2006 to 2009, he was doing research work on medical imaging and cardiac flow analysis. He has over ten years of medical imaging, computational hemodynamics, and simulation modeling experience. Since 2009, he has been involved in computational fluid and solid dynamics. In 2013, he was involved in the discrete element method with applications in biological systems with The University of Western Australia. He was the originator of the spatial game moment concept, in which multi-objective multi-constraint combinatorial optimization problems can be treated as decision-making problems in the game theoretical sense, and solved with high efficiency. In addition, he is the first author of the book Methods in Research and Development of Biomedical Devices (World Scientific, 2013), and a co-author of a second book Computational Hemodynamics—Theory, Modeling and Applications (Springer, 2015). His publications now span a diverse range of topics in the medical science and engineering field. He has served as associate editors and guest editors for journals in the area of biomedical engineering.

Giancarlo Fortino is currently Full Professor of Computer Engineering at the Department of Informatics, Modeling, Electronics and Systems (DIMES) of the University of Calabria (Unical), Rende (CS), Italy. In 2013 he took the Italian Scientific Habilitation for Full Professorship. He received a Laurea Degree (BSc+MSc) in Computer Engineering and a Ph.D. in Computer Engineering from the University of Calabria, Italy, in 1995 and 2000, respectively. He has been a *Research Scholar* at the International Computer Science Institute (ICSI), Berkeley (CA), USA, in 1997 and 1999, and *Visiting Professor* at Queensland University of Technology (QUT), Brisbane, Australia, in 2009. He was also Assistant Professor (2001–2006) and Associate Professor (2006–2018) at University of Calabria. He was nominated *Guest Professor* in Computer Engineering of Wuhan University of Technology on April, 18 2012 (the term of appointment is three years) and in April 2015 he was nominated Adjunct Full Professor of Computer Engineering in the framework of High-End Foreign Experts in China (the term of appointment is three years). He has been nominated Adjunct Senior Research Fellow at ICAR-CNR (Italian National Research Council), for the year 2015. In 2017 he was also nominated High-end Expert at HUST (Huazhong University of Science and Technology), China, and recently he won a CAS PIFI grant at SIAT in Shenzhen (China). His research interests include distributed computing and networks, agent systems, agent oriented software engineering, body area networks and wearable computing, wireless sensor networks, Internet of Things (including social and trust mechanism), multimedia streaming content distribution networks, GRID/Cloud computing. He has been involved in research and development projects in the research/technical areas of Internet of Things, wireless sensor networks, intelligent and mobile agents, and content delivery networks. He gained funding for about 3MEuro. Currently, he is the deputy coordinator and STPM of the EU-funded H2020 project INTER-IoT (Interoperability of IoT Platforms). He is author of many papers (400+) in international journals (more than 100 papers in Top-level Journals including IEEE, ACM and Elsevier), conferences (majority IEEE conferences) and books (edited by Springer and IEEE Press). He is High-Impact Researcher according to Scopus in the last 5 years 2013-18: he is ranked n. 56 in the World in the Computer Science field and N.1 in the research theme "Middleware, Internet and Sensor Networks". He is IEEE Senior Member (member of the IEEE Computer, Communications, SMC, and IoT societies), co-chair of the IEEE SMC TC on "Interactive and Wearable Computing and Devices", member of the IEEE Press Editorial Board, and also member of ACM. He is notably Member of the BoG of the IEEE SMC Society (term 2018-2020) and the IEEE SMC Society Italian Chapter Chair.



Derek Abbott was born in London, U.K., in 1960. He received the B.Sc. degree (Hons.) in physics from Loughborough University, Leicestershire, U.K., in 1982, and the Ph.D. degree in electrical and electronic engineering from The University of Adelaide, Adelaide, SA, Australia, in 1995, under the supervision of K. Eshraghian and B. R. Davis. From 1978 to 1986, he was a Research Engineer with the GEC Hirst Research Centre, London. From 1986 to 1987, he was a VLSI Design Engineer with Austek Microsystems, Australia. Since 1987, he has been with The University of Adelaide, where he is currently a Full Professor with the School of Electrical and Electronic Engineering. He coedited Quantum Aspects of Life (London, U.K.:

Imperial College Press, 2008), co-authored Stochastic Resonance (Cambridge, U.K.: Cambridge University Press, 2012), and co-authored Terahertz Imaging for Biomedical Applications (New York, NY, USA: SpringerVerlag, 2012). He holds over 800 publications/patents and has been an invited speaker at over 100 institutions. His research interests include the area of multidisciplinary physics, electronic engineering applied to complex systems, stochastics, game theory, photonics, biomedical engineering, and computational neuroscience.

Prof. Abbott is a fellow of the Institute of Physics. He received the number of awards, including the South Australian Tall Poppy Award for Science in 2004, the Premier's SA Great Award in Science and Technology for outstanding contributions to South Australia in 2004, and an Australian Research Council Future Fellowship in 2012. With his colleagues, he received the IEEE SENSORS Journal Best Paper Award in 2014. He received the David Dewhurst Medal for biomedical engineering in 2015. He has served as an Editor and/or Guest Editor for a number of journals, including the IEEE JOURNAL OF SOLID-STATE CIRCUITS, the Journal of Optics B, Microelectronics Journal, Chaos, Smart Structures and Materials, Fluctuation and Noise Letters, PLOS One, the PROCEEDINGS OF THE IEEE, the IEEE PHOTONICS JOURNAL. He is currently on the Editorial Boards of Scientific Reports (Nature), Royal Society Open Science, the IEEE Access, and Frontiers in Physics.