Deep Learning for Health Informatics

Daniele Ravì, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo,
and Guang-Zhong Yang, *Fellow, IEEE*

Abstract—With a massive influx of multimodality data, the role of data analytics in health informatics has grown rapidly in the last decade. This has also prompted increasing interests in the generation of analytical, data driven models based on machine learning in health informatics. Deep learning, a technique with its foundation in artificial neural networks, is emerging in recent years as a powerful tool for machine learning, promising to reshape the future of artificial intelligence. Rapid improvements in computational power, fast data storage, and parallelization have also contributed to the rapid uptake of the technology in addition to its predictive power and ability to generate automatically optimized high-level features and semantic interpretation from the input data. This article presents a comprehensive up-to-date review of research employing deep learning in health informatics, providing a critical analysis of the relative merit, and potential pitfalls of the technique as well as its future outlook. The paper mainly focuses on key applications of deep learning in the fields of translational bioinformatics, medical imaging, pervasive sensing, medical informatics, and public health.

Index Terms—Bioinformatics, deep learning, health informatics, machine learning, medical imaging, public health, wearable devices.
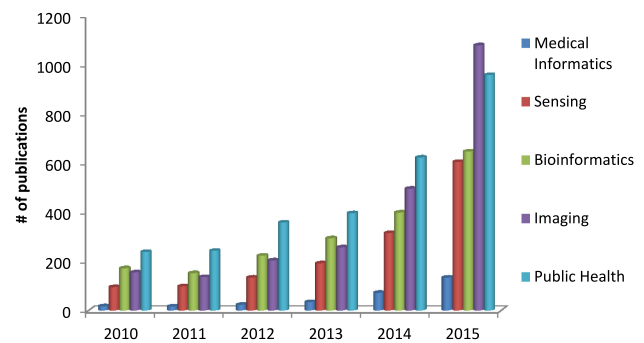
Fig. 1. Distribution of published papers that use deep learning in subareas of health informatics. Publication statistics are obtained from Google Scholar; the search phrase is defined as the subfield name with the exact phrase *deep learning* and at least one of *medical* or *health* appearing, e.g., "public health" "deep learning" medical OR health.

I. INTRODUCTION

DEEP learning has in recent years set an exciting new trend in machine learning. The theoretical foundations of deep learning are well rooted in the classical neural network (NN) literature. But different to more traditional use of NNs, deep learning accounts for the use of many hidden neurons and layers—typically more than two—as an architectural advantage combined with new training paradigms. While resorting to many neurons allows an extensive coverage of the raw data at hand, the layer-by-layer pipeline of nonlinear combination of their outputs generates a lower dimensional projection of the input space. Every lower-dimensional projection corresponds to a higher perceptual level. Provided that the network is optimally weighted, it results in an effective high-level abstraction of the raw data or images. This high level of abstraction renders

an automatic feature set, which otherwise would have required hand-crafted or bespoke features.

In domains such as health informatics, the generation of this automatic feature set without human intervention has many advantages. For instance, in medical imaging, it can generate features that are more sophisticated and difficult to elaborate in descriptive means. Implicit features could determine fibroids and polyps [1], and characterize irregularities in tissue morphology such as tumors [2]. In translational bioinformatics, such features may also determine nucleotide sequences that could bind a DNA or RNA strand to a protein [3]. Fig. 1 outlines a rapid surge of interest in deep learning in recent years in terms of the number of papers published in sub-fields in health informatics including bioinformatics, medical imaging, pervasive sensing, medical informatics, and public health.

Among various methodological variants of deep learning, several architectures stand out in popularity. Fig. 2 depicts the number of publications by deep learning method since 2010. In particular, Convolutional Neural Networks (CNNs) have had the greatest impact within the field of health informatics. Its architecture can be defined as an interleaved set of feed-forward layers implementing convolutional filters followed by reduction, rectification or pooling layers. Each layer in the network originates a high-level abstract feature. This biologically-inspired architecture resembles the procedure in which the visual cortex assimilates visual information in the form of receptive fields. Other plausible architectures for deep learning include those grounded in compositions of restricted Boltzmann machines (RBMs) such as deep belief networks (DBNs), stacked Autoencoders functioning as deep Autoencoders, extending artificial
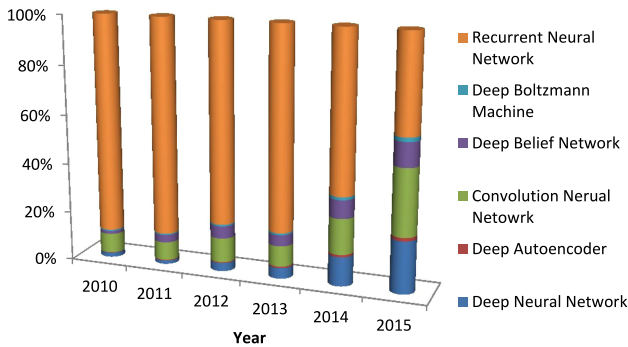
Fig. 2. Percentage of most used deep learning methods in health informatics. Learning method statistics are also obtained from Google Scholar by using the method name with at least one of *medical* or *health* as the search phrase.

NNs with many layers as deep neural networks (DNNs), or with directed cycles as recurrent neural networks (RNNs). Latest advances in Graphics Processing Units (GPUs) have also had a significant impact on the practical uptake and acceleration of deep learning. In fact, many of the theoretical ideas behind deep learning were proposed during the pre-GPU era, although they have started to gain prominence in the last few years. Deep learning architectures such as CNNs can be highly parallelized by transferring most common algebraic operations with dense matrices such as matrix products and convolutions to the GPU.

Thus far, a plethora of experimental works have implemented deep learning models for heath informatics, reaching similar performance or in many cases exceeding that of alternative techniques. Nevertheless, the application of deep learning to health informatics raises a number of challenges that need to be resolved. For example, training a deep architecture requires an extensive amount of labeled data, which in the healthcare domain can be difficult to achieve. In addition, deep learning requires extensive computational resources, without which training could become excessively time-consuming. Attaining an optimal definition of the network's free parameters can become a particularly laborious task to solve. Eventually, deep learning models can be affected by convergence issues as well as overfitting, hence supplementary learning strategies are required to address these problems [4].

In the following sections of this review, we examine recent health informatics studies that employ deep learning to discuss its relative strength and potential pitfalls. Furthermore, their schemas and operational frameworks are described in detail to elucidate their practical implementations, as well as expected performance.

II. FROM PERCEPTRON TO DEEP LEARNING

Perceptron is a bio-inspired algorithm for binary classification and it is one of the earliest NNs proposed [19]. It mathematically formalizes how a biological neuron works. It has been realized that the brain processes information through billions of these interconnected neurons. Each neuron is stimulated by the injection of currents from the interconnected neurons and an action potential is generated when the voltage exceeds a limit. These action potentials allow neurons to excite or inhibit other

neurons, and through these networked neural activities, the biological network can encode, process, and transmit information. Biological NNs have the capacity to modify themselves, create new neural connections, and learn according to the stimulation characteristics. Perceptrons, which consist of an input layer directly connected to an output node, emulate this biochemical process through an activation function (also referred to as a transfer function) and a few weights. Specifically, it can learn to classify linearly separable patterns by adjusting these weights accordingly.

To solve more complex problems, NNs with one or more hidden layers of Perceptrons have been introduced [20]. To train these NNs, many stages or epochs are usually performed where each time the network is presented with a new input sample and the weights of each neuron are adjusted based on a learning process called delta rule. The delta rule is used by the most common class of supervised NNs during the training and is usually implemented by exploiting the back-propagation routine [21]. Specifically, without any prior knowledge, random values are assigned to the network weights. Through an iterative training process, the network weights are adjusted to minimize the difference between the network outputs and the desired outputs. The most common iterative training method uses the gradient descent method where the network is optimized to find the minimum along the error surface. The method requires the activation functions to be always differentiable.

Adding more hidden layers to the network allows a deep architecture to be built that can express more complex hypotheses as the hidden layers capture the nonlinear relationships. These NNs are known as DNNs. Training of DNNs is not trivial because once the errors are back-propagated to the first few layers, they become negligible (vanishing of the gradient), thus failing the learning process. Although more advanced variants of back-propagation [22] can solve this problem, they still result in a very slow learning process.

Deep learning has provided new sophisticated approaches to train DNN architectures. In general, DNNs can be trained with unsupervised and supervised learning methodologies. In supervised learning, labeled data are used to train the DNNs and learn the weights that minimize the error to predict a target value for classification or regression, whereas in unsupervised learning, the training is performed without requiring labeled data. Unsupervised learning is usually used for clustering, feature extraction or dimensionality reduction. For some applications it is common to combine an initial training procedure of the DNN with an unsupervised learning step to extract the most relevant features and then use those features for classification by exploiting a supervised learning step. For more general background information related to the theory of machine learning, the reader can refer to the works in [23]–[25] where common training problems, such as overfitting, model interpretation and generalization, are explained in detail. These considerations must be taken into account when deep learning frameworks are used.

For many years, hardware limitations have made DNNs impractical due to high computational demands for both training and processing, especially for applications that require real-time processing. Recently, advances in hardware and thanks to the possibility of parallelization through GPU acceleration,

cloud computing and multicore processing, these limitations have been partially overcome and have enabled DNNs to be recognized as a significant breakthrough in artificial intelligence. Thus far, several DNNs architectures have been introduced in literature and Table I briefly describes the pros and cons of the commonly used deep learning approaches in the field of health informatics. In Table II are instead described the main features of popular software packages that provide deep learning implementation. Finally, Table III summarizes the different applications in the five areas of health informatics considered in this paper.

A. Autoencoders and Deep Autoencoders

Recent studies have shown that there are no universally hand-engineered features that always work on different datasets. Features extracted using data driven learning can generally be more accurate. An Autoencoder is a NN designed exactly for this purpose. Specifically, an Autoencoder has the same number of input and output nodes, as shown in Fig. 3(a), and it is trained to recreate the input vector rather than to assign a class label to it. The method is therefore unsupervised. Usually, the number of hidden units is smaller than the input/output layers, which achieve encoding of the data in a lower dimensional space and extract the most discriminative features. If the input data is of high dimensionality, a single hidden layer of an Autoencoder may not be sufficient to represent all the data. Alternatively, many Autoencoders can be stacked on top of each other to create a deep Autoencoder architecture [5]. Deep Autoencoder structures also face the problem of vanishing gradients during training. In this case, the network learns to reconstruct the average of all the training data. A common solution to this problem is to initialize the weights so that the network starts with a good approximation of the final configuration. Finding these initial weights is referred to as pretraining and is usually achieved by training each layer separately in a greedy fashion. After pretraining, the standard back-propagation can be used to fine-tune the parameters. Many variations of Autoencoder have been proposed to make the learned representations more robust or stable against small variations of the input pattern. For example, the sparse autoencoder [6] that forces the representation to be sparse is usually used to make the classes more separable. Another variation, called denoising autoencoder, was proposed by Vincent et al. [7], where in order to increase the robustness of the model, the method recreates the input introducing some noise to the patterns, thus, forcing the model to capture just the structure of the input. A similar idea was implemented in contractive autoencoder, proposed by Rifai et al. [8], but instead of injecting noise to corrupt the training set, it adds an analytic contractive penalty to the error function. Finally, the convolutional autoencoder [9] shares weights between nodes to preserve spatial locality and process two-dimensional (2-D) patterns (i.e., images) efficiently.

B. Recurrent Neural Network

RNN [13] is a NN that contains hidden units capable of analyzing streams of data. This is important in several
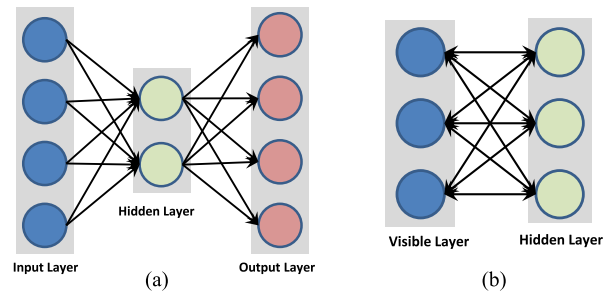

Fig. 3. Schematic illustration of simple NNs without deep structures. (a) Autoencoder. (b) Restricted Boltzmann machine.

applications where the output depends on the previous computations, such as the analysis of text, speech, and DNA sequences. The RNN is usually fed with training samples that have strong inter-dependencies and a meaningful representation to maintain information about what happened in all the previous time steps. The outcome obtained by the network at time $t - 1$ affects the choice at time t. In this way, RNNs exploit two sources of input, the present and the recent past, to provide the output of the new data. For this reason, it is often said that RNNs have memory. Although the RNN is a simple and powerful model, it also suffers from the vanishing gradient and exploding gradient problems as described in Bengio et al. [26]. A variation of RNN called long short-term memory units (LSTMs), was proposed in [27] to solve the problem of the vanishing gradient generated by long input sequences. Specifically, LSTM is particularly suitable for applications where there are very long time lags of unknown sizes between important events. To do so, LSTMs exploit new sources of information so that data can be stored in, written to, or read from a node at each step. During the training, the network learns what to store and when to allow reading/writing in order to minimize the classification errors.

Unlike other types of DNNs, which uses different weights at each layer, a RNN or a LSTM shares the same weights across all steps. This greatly reduces the total number of parameters that the network needs to learn. RNNs have shown great successes in many natural language processing tasks such as language modeling, bioinformatics, speech recognition, and generating image description.

C. RBM-Based Technique

A RBM was first proposed in [37] and is a variant of the Boltzmann machine, which is a type of stochastic NN. These networks are modeled by using stochastic units with a specific distribution (for example Gaussian). Learning procedure involves several steps called Gibbs sampling, which gradually adjust the weights to minimize the reconstruction error. Such NNs are useful if it is required to model probabilistic relationships between variables.

Bayesian networks [38], [39] are a particular case of network with stochastic unit referred as probabilistic graphical model that characterizes the conditional independence between variables in the form of a directed acyclic graph. In an RBM, the visible
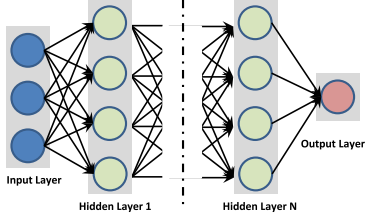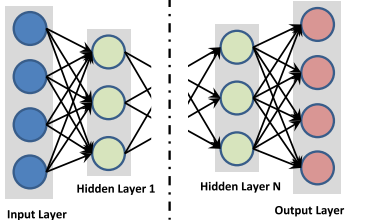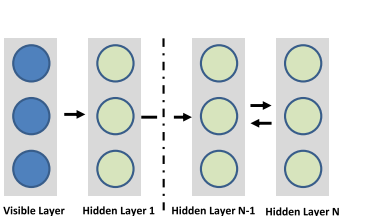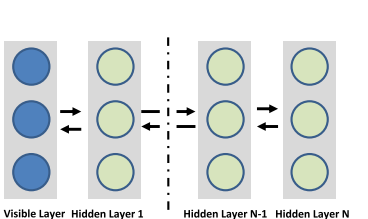
TABLE I
DIFFERENT DEEP LEARNING ARCHITECTURES

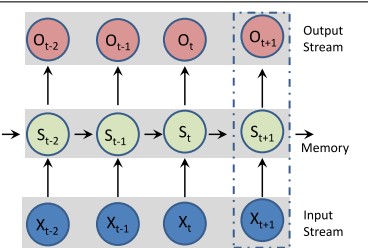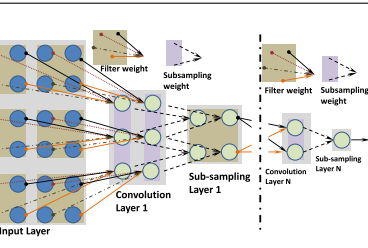| Architecture	Description	Key Points
	Deep Neural Network • General deep framework usually used for classification or regression • Made of many hidden layers (more than 2) • Allows complex (non-linear) hypotheses to be expressed	**Pros:** • Widely used with successes in many areas **Cons:** • Training is not trivial because once the errors are back-propagated to the first few layers they become minuscule • The learning process can be very slow
	Deep Autoencoder • Proposed in [5] and is mainly designed for feature extraction or dimensionality reduction • Has the same number of input and output nodes • Aims to recreate the input vector • Unsupervised learning method	**Pros:** • Does not require labelled data for training • Many variations have been proposed to make the representation more robust: Sparse AutEnc. [6], Denoising AutEnc. [7], Contractive AutEnc. [8], Convolutional AutEnc. [9] **Cons:** • Requires a pre-training stage • Training can also suffer from vanishing of the errors
	Deep Belief Network • Proposed in [10] is a composition of RBM where each sub-network's hidden layer serves as the visible layer for the next • Has undirected connections just at the top two layers • Allows unsupervised and supervised training of the network	**Pros:** • Proposes a layer-by-layer greedy learning strategy to initialize the network • Inferences tractable maximizing the likelihood directly **Cons:** • Training procedure is computationally expensive due to the initialization process and sampling
	Deep Boltzmann Machine • Proposed in [11] is another approach based on the Boltzmann family • Possesses undirected connections (conditionally independent) between all layers of the network • Uses a stochastic maximum likelihood [12] algorithm to maximize the lower bound of the likelihood	**Pros:** • Incorporates top-down feedback for a more robust inferences with ambiguous inputs **Cons:** • Time complexity for the inference is higher than DBN • Optimization of the parameters is not practical for large datasets
	Recurrent Neural Network • Proposed in [13] is a NN capable of analyzing stream of data • Useful in applications where the output depends on the previous computations • Shares the same weights across all steps	**Pros:** • Can memorize sequential events • Can model time dependencies • Has shown great success in many Natural Language Processing applications **Cons:** • Learning issues are frequent due to the vanishing gradient and exploding gradient problems
	Convolutional Neural Network • Proposed in [14], it is well suited for 2D data such as images • Every hidden convolutional filter transforms its input to a 3D output volume of neuron activations • Inspired by the neurobiological model of the visual cortex [15]	**Pros:** • Few neuron connections required with respect to a typical NN • Many variants have been proposed: AlexNet [16], Clarifai [17], and GoogLeNet [18] **Cons:** • It may require many layers to find an entire hierarchy of visual features • It usually requires a large dataset of labelled images

TABLE II
POPULAR SOFTWARE PACKAGES THAT PROVIDE DNNS IMPLEMENTATION

Name	Creator	License	Platform	Interface	OpenMP support	Supported techniques			Cloud computing
						RNN	CNN	DBN	
Caffe [28]	Berkeley Center	FreeBSD	Linux, Win, OSX, Andr.	C++, Python, MATLAB	✗	√	√	✗	✗
CNTK [29]	Microsoft	MIT	Linux, Win	Command line	√	√	√	✗	✗
Deeplearning4jK [30]	Skymind	Apache 2.0	Linux, Win, OSX, Andr.	Java, Scala, Clojure	√	√	√	√	✗
Wolfram Math. [31]	Wolfram Research	Proprietary	Linux, Win, OSX, Cloud	Java, C++	✗	✗	√	√	√
TensorFlow [32]	Google	Apache 2.0	Linux, OSX	Python	✗	√	√	√	✗
Theano [33]	Université de Montréal	BSD	Cross-platform	Python	√	√	√	√	✗
Torch [34]	Ronan Collobert *et al.*	BSD	Linux, Win, OSX, Andr., iOS	Lua, LuaJIT, C	√	√	√	√	✗
Keras [35]	Franois Chollet	MIT license	Linux, Win, OSX	Python	✗	√	√	√	✗
Neon [36]	Nervana Systems	Apache 2.0	OSX, Linux	Python	√	√	√	√	√

TABLE III
SUMMARY OF THE DIFFERENT DEEP LEARNING METHODS BY AREAS AND APPLICATIONS IN HEALTH INFORMATICS

Area	Applications	Input data	Base Method	Reference
Bioinformatics	Cancer diagnosis	Gene expression	Deep Autoencoders	[2]
	Gene selection/classification	MicroRNA	Deep Belief Network	[46], [47]
	Gene variants	Microarray data	Deep Neural Network	[48]
	Drug design	Molecule compounds	Deep Neural Network	[49]
	Compound-Protein interaction	Protein structures	Deep Belief Network	[50]
	RNA binding protein	Molecule compounds		
	DNA methylation	Genes/RNA/DNA sequences	Deep Neural Network	[51], [52]
Medical Imaging	3D brain reconstruction	MRI/fMRI	Deep Autoencoders	[53], [54]
	Neural cells classification		Convolutional Neural Network	[55]–[59]
	Brain tissues classification	Fundus images	Deep Belief Network	[60], [61]
	Alzheimer/MCI diagnosis	PET scans	Deep Near Network	[62]
	Tissue classification	MRI/CT Images	Convolutional Deep Belief Network	[63], [64]
	Organ segmentation	Endoscopy images	Convolutional Neural Network	[65]–[76]
	Cell clustering	Microscopy	Deep Autoencoder	[66], [77]
	Hemorrhage detection	Fundus Images	Group Method of Data Handling	[78]–[81]
	Tumour detection	X-ray images	Deep Neural Network	[82]–[85]
		Hyperspectral images		
Pervasive Sensing	Anomaly detection	EEG	Deep Belief Network	[86]–[89]
	Biological parameters monitoring	ECG		
		Implantable device		
	Human activity recognition	Video	Convolutional Neural Network	[90]–[93]
		Wearable device	Deep Belief Network	[94], [95]
			Deep Neural Network	[96]
	Hand gesture recognition	Depth camera	Convolutional Neural Network	[97]
	Obstacle detection	RGB-D camera	Deep Belief Network	[98]
	Sign language recognition	Real-Sense camera		
	Food intake	Wearable device	Convolutional Neural Network	[99]
	Energy expenditure	RGB Image	Deep Neural Network	[100]
		Mobile device		
Medical Informatics	Prediction of disease	Electronic health records	Deep Autoencoders	[101], [102]
	Human behaviour monitoring	Big medical dataset	Deep Belief Network	[103], [104]
	Data mining	Blood/Lab tests	Convolutional Neural Network	[105]
			Recurrent Neural Network	[101], [106]
			Convolutional Deep Belief Network	[107]
			Deep Neural Network	[108], [109]
Public Health	Predicting demographic info	Social media data	Deep Autoencoders	[110]
	Lifestyle diseases	Mobile phone metadata	Deep Belief Network	[111], [112]
	Infectious disease epidemics	Geo-tagged images	Convolutional Neural Network	[113]
	Air pollutant prediction	Text messages	Deep Neural Network	[114]–[117]

and hidden units are restricted to form a bipartite graph that allows implementation of more efficient training algorithms. Another important characteristics is that RBMs have undirected nodes, which implies that values can be propagated in both the directions as shown in Fig. 3(b).

Contrastive divergence [40] (CD) algorithm is a common method used to train an RBM. CD is an unsupervised learning algorithm, which consists of two phases that can be referred to as positive and negative phases. During the positive phase the network configuration is modified to replicate the training set,

whereas during the negative phase it attempts to recreate the data based on the current network configuration.

A beneficial property of RBM is that the conditional distribution over the hidden units factorizes given the visible units. This makes inferences tractable since the RBM feature representation is taken to be a set of posterior marginal obtained by directly maximizing the likelihood. Utilizing RBM as learning modules, two main deep learning frameworks have been proposed in literature: the DBN and the deep Boltzmann machine (DBM).

1) Deep Belief Network: Proposed in [10], a DBN can be viewed as a composition of RBMs where each subnetwork's hidden layer is connected to the visible layer of the next RBM. DBNs have undirected connections only at the top two layers and directed connections to the lower layers. The initialization of a DBN is obtained through an efficient layer-by-layer greedy learning strategy using unsupervised learning and is then fine-tuned based on the target outputs.

2) Deep Boltzmann Machines: Proposed in [11], a DBM is another DNN variant based on the Boltzmann family. The main difference with DBN is that the former possesses undirected connections (conditionally independent) between all layers of the network. In this case, calculating the posterior distribution over the hidden units given the visible units cannot be achieved by directly maximizing the likelihood due to interactions between the hidden units. For this reason, to train a DBM, a stochastic maximum likelihood [12] based algorithm is usually used to maximize the lower bound of the likelihood. Same as for DBNs, a greedy layer-wise training strategy is also performed when pretraining the DBM network. The main disadvantage of a DBM is the time complexity required for the inference that is considerably higher with respect to the DBN, and that makes the optimization of the parameters not practical for big training set [41].

D. Convolutional Neural Networks

In general, all the DNNs presented so far cannot scale well with multidimensional input that has locally correlated data, such as an image. The main problem is that the number of nodes and the number of parameters that they have to train could be huge, and therefore, they are not practical. CNNs have been proposed in [14] to analyze imagery data. The name of these networks comes from the convolution operator that is an easy way to perform complex operations using convolution filter. CNN does not use predefined kernels, but instead learns locally connected neurons that represent data-specific kernels. Since these filters are applied repeatedly to the entire image, the resulting connectivity looks like a series of overlapping receptive fields. The main advantage of a CNN is that during back-propagation, the network has to adjust a number of parameters equal to a single instance of the filter which drastically reduces the connections from the typical NN architecture. The concept of CNN is largely inspired by the neurobiological model of the visual cortex [15]. The visual cortex is known to consist of maps of local receptive fields that decrease in granularity as the cortex moves anteriorly. This process can be briefly summarized as follows:
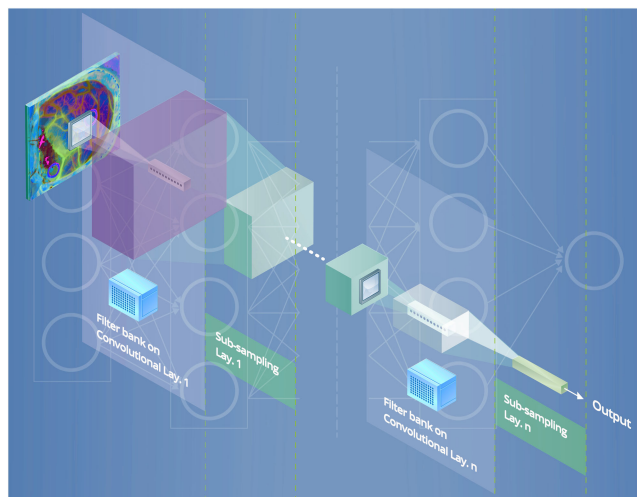

Fig. 4. Basic architecture of CNN which consists in several layers of convolution and subsampling to efficiently process images.
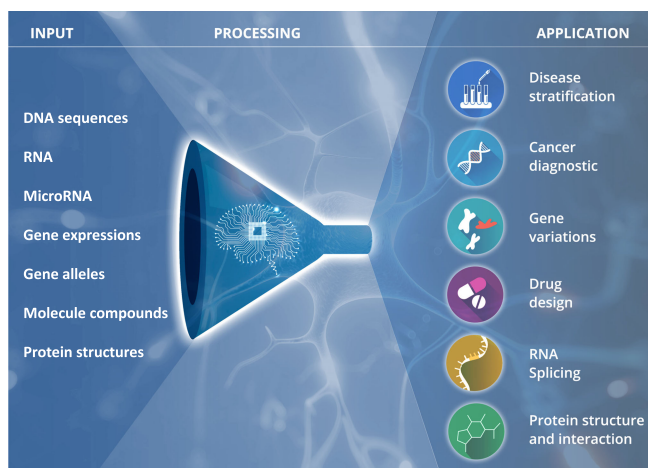
Fig. 5. Overview of the different inputs and applications in biomedical and health informatics.

1) The input image is convolved using several small filters.
2) The output at Step 1 is subsampled.
3) The output at Step 2 is considered the new input and the convolution and subsampling processes are repeated until high level features can be extracted.

According to the aforementioned schema, a typical CNN configuration consists of a sequence of convolution and subsample layers as illustrated in Fig. 4. After the last subsampling layer, a CNN usually adopts several fully-connected layers with the aim of converting the 2-D feature maps into a 1-D vector to allow final classification. Fully-connected layers can be considered like traditional NNs and they contain about 90% of the parameters of the entire CNN, which increases the effort required for training considerably. A common solution for solving this problem is to decrease the connections in these layers with a sparsely connected architecture. To this end, many configurations and variants have been proposed in literature and some of the most popular CNNs at the moment are: AlexNet [16], Clarifai [17], VGG [42], and GoogLeNet [18].

A more recent deep learning approach is known as convolutional deep belief networks (CDBN) [43]. CDBN maintains structures that are very similar to a CNN, but is trained similarly to a DBN. Therefore, it exploits the advantages of CNN whilst making use of pretraining to initialize efficiently the network as a DBN does.

E. Software/Hardware Implementations

Table II lists the most popular software packages that allow implementation of customized deep learning methodologies based on the approaches described so far. All the software listed in the table can exploit CUDA/Nvidia support to improve performance using GPU acceleration. Adding to the growing trend of proprietary deep learning frameworks being turned into open source projects, some companies, such as Wolfram Mathematica [31] and Nervana Systems [36], have decided to provide a cloud based services that allow researchers to speed-up the training process. New GPU acceleration hardware includes purpose-built microprocessors for deep learning, such as the Nvidia DGX-1 [44]. Other possible future solutions are neuromorphic electronic systems that are usually used in computational neuroscience simulations. These later hardware architectures intend to implement artificial neurons and synapses in a chip. Some current hardware designs are IBM TrueNorth, SpiNNaker [45], NuPIC, and Intel Curie.

III. APPLICATIONS

A. Translational Bioinformatics

Bioinformatics aims to investigate and understand biological processes at a molecular level. The human genome project has made available a vast amount of unexplored data and allowed the development of new hypotheses of how genes and environmental factors interact together to create proteins [118], [119]. Further advances in biotechnology have helped reduce the cost of genome sequencing and steered the focus on prognostic, diagnostic and treatment of diseases by analyzing genes and proteins. This can be illustrated by the fact that sequencing the first human genome cost billions of dollars, whereas today it is affordable [45]. Further motivated by P4 (predictive, personalized, preventive, participatory) medicine [120], bioinformatics aims to predict and prevent diseases by involving patients in the development of more efficient and personalized treatments.

The application of machine learning in bioinformatics (Fig. 5) can be divided into three areas: prediction of biological processes, prevention of diseases and personalized treatment. These areas are found in genomics, pharmacogenomics and epigenomics. Genomics explores the function and information structures encoded in the DNA sequences of a living cell [121]: it analyzes genes or alleles responsible for the creation of protein sequences and the expression of phenotypes. A goal of genomics is to identify gene alleles and environmental factors that contribute to diseases such as cancer. Identification of these genes can enable the design of targeted therapies [121]. Pharmacogenomics evaluates variations in an individual's drug response to treatment brought about by differences in genes. It aims to design more efficient drugs for personalized treatment whilst

reducing side effects. Finally, epigenomics aims to investigate protein interactions and understand higher level processes, such as transcriptome (mRNA count), proteome, and metabolome, which lead to modification in the gene's expression. Understanding how environmental factors affect protein formation and their interactions is a goal of epigenomics.

1) Genetic Variants: splicing and alternative splicing code. Genetic variant aims to predict human splicing code in different tissues and understand how gene expression changes according to genetic variations. Alternative splicing code is the process from which different transcripts are generated from one gene. Prediction of splicing patterns is crucial to better understand genes variations, phenotypes consequences and possible drug effect variations. Genetic variances play a significant role in the expression of several diseases and disorders, such as autism, spinal muscular atrophy, and hereditary colorectal cancer. Therefore, understanding genetic variants can be a key to provide early diagnosis.

2) Protein–Protein and Compound-Protein Interactions (CPI): Quantitative structure activity relationship (QSAR) aims to predict the protein–protein interaction normally based on structural molecular information. CPI aims to predict the interaction between a given compound and protein. Protein–protein and protein-compound interactions are important in virtual screening for drug discovery: they help identifying new compounds, toxic substances, and provide significant interpretation on how a drug will affect any type of cell, targeted or not. Specifically to epigenomics, QSAR and CPI help modeling the RNA protein binding.

3) DNA Methylation: DNA methylation states are part of a process that changes the DNA expression without changing the DNA sequence itself. This can be brought about by a wide range of reasons, such as chromosome instability, transcription or translation errors, cell differentiation or cancer progression.

The datasets are usually high dimensional, heterogeneous, and sometimes unbalanced. The conventional workflow includes data preprocessing/cleaning, feature extraction, model fitting, and evaluation [122]. These methods do not operate on the sequence data directly but they require domain knowledge. For example, the ChEMBL database, used in pharmacogenomics, has millions of compounds and compound descriptors associated with a large database of drug targets [45]. Such databases encode molecular "fingerprints" and are major sources of information in drug discovery applications. Traditional machine learning approaches have been successful, mostly because the complexity of molecular interactions was reduced by only investigating one or two dimension of the molecule structure in the feature descriptors. Reducing design complexity inevitably leads to ignore some relevant but uncaptured aspects of the molecular structures [123], [124]. However, Zhang et al. [50] used deep learning to model structural features for RNA binding protein prediction and showed that using the RNA tertiary structural profile can improve outcomes.

Extracting biomarkers or alleles of genes responsible for a specific disorder is very challenging as it requires a great amount of data from a large diversified cohort. The markers should be present—if possible at different concentration levels throughout the disorder's evolution and patient's treatment—with a direct

explanation on the phenotype changes due to the disease [125]. One approach accounting for sequence variation which limits the number of required subjects is to split the sequence into windows centered on the investigated trait. Although this results in thousands of training examples of molecular traits even from just one genome, a large scale of DNA sequences and interactions mediated by various distant regulatory factors should be used [122].

The ability of deep learning to abstract large, complex, and unstructured data offers a powerful way of analyzing heterogeneous data such as gene alleles, proteins occurrences, and environmental factors [126]. Their contribution to bioinformatics has been reviewed in several related areas [45], [121], [122], [124], [126]–[129]. In deep learning approaches, feature extraction and model fitting takes place in a unified step. Multilayer feature representation can capture nonlinear dependencies at multiple scales of transcriptional and epigenetic interactions and can model molecular structure and properties in a data-driven way. These nonlinear features are invariant to small input changes which results in eliminating noise and increasing the robustness of the technique.

Several works have demonstrated that deep learning features outperformed methods relying on visual descriptors in the recognition and classification of cancer cells. For example, Fakoor *et al.* [2] proposed an autoencoder architecture based on gene expression data from different types of cancer and the same microarray dataset to detect and classify cancer. Ibrahim *et al.* [46] proposed a DBN with an active learning approach to find features in genes and microRNA that resulted in the best classification performance of various cancer diseases such as hepatocellular carcinoma, lung cancer and breast cancer. For breast cancer genetic detection, Khademi *et al.* [47] overcame missing attributes and noise by combining a DBN and Bayesian network to extract features from microarray data. Deep learning approaches have also outperformed SVM in predicting splicing code and understanding how gene expression changes by genetic variants [48], [130]. Angermueller *et al.* [52] used DNN to predict DNA methylation states from DNA sequences and incomplete methylation profiles. After applying to 32 embryonic mice stem cells, the baseline model was compared with the results. This method can be used for genome-wide downstream analyses.

Deep learning not only outperforms conventional approaches but also opens the door to more efficient methods to be developed. Kearnes *et al.* [123] described how deep learning based on graph convolutions can encode molecular structural features, physical properties, and activities in other assays. This allows a rich representation of possible interactions beyond the molecular structural information encoded in standard databases. Similarly, multitask DNNs provides an intuitive model of correlation between molecule compounds and targets because information can be shared among different nodes. This increases robustness, reduces chances to miss information, and usually outperforms other methods that process large datasets [49].

Deep learning has rapidly been adopted in the field of bioinformatics due to several open source packages. However, there are no standard methods of choosing model architectures and

their use require expertise in computer science and biology. Therefore, the question of integrating the software development and the data has been raised [127]. Also, deep learning approaches do not include a standard way of establishing statistical significance, which is a limitation for future result comparisons. Therefore, conventional methods offer some advantages, especially in the case of small datasets. Although DNNs scale better to large datasets, the computational cost is high, resulting in the specific necessity of chips for massive parallel processing in order to deal with the increased complexity [45].

B. Deep Learning for Medical Imaging

Automatic medical imaging analysis is crucial to modern medicine. Diagnosis based on the interpretation of images can be highly subjective. Computer-aided diagnosis (CAD) can provide an objective assessment of the underlying disease processes. Modeling of disease progression, common in several neurological conditions, such as Alzheimer's, multiple sclerosis, and stroke, requires analysis of brain scans based on multimodal data and detailed maps of brain regions.

In recent years, CNNs have been adapted rapidly by the medical imaging research community because of their outstanding performance demonstrated in computer vision and their ability to be parallelized with GPUs. The fact that CNNs in medical imaging have yielded promising results have also been highlighted in a recent survey of CNN approaches in brain pathology segmentation [58] and in an editorial of deep learning techniques in computer aided detection, segmentation, and shape analysis [76].

Among the biggest challenges in CAD are the differences in shape and intensity of tumors/lesions and the variations in imaging protocol even within the same imaging modality. In several cases, the intensity range of pathological tissue may overlap with that of healthy samples. Furthermore, Rician noise, non-isotropic resolution, and bias field effects in magnetic resonance images (MRI) cannot be handled automatically using simpler machine learning approaches. To deal with this data complexity, hand-designed features are extracted and conventional machine learning approaches are trained to classify them in a completely separate step.

Deep learning provides the possibility to automate and merge the extraction of relevant features with the classification procedure [55], [65]. CNNs inherently learn a hierarchy of increasingly more complex features and, thus, they can operate directly on a patch of images centered on the abnormal tissue. Example applications of CNNs in medical imaging include the classification of interstitial lung diseases based on computed tomography (CT) images [70], the classification of tuberculosis manifestation based on X-ray images [71], the classification of neural progenitor cells from somatic cell source [57], the detection of haemorrhages in color fundus images [69] and the organ or body-part-specific anatomical classification of CT images [68]. A body-part recognition system is also presented in Yan *et al.* [75]. A multistage deep learning framework based on CNNs extracts both the patches with the most as well as least discriminative local patches in the pretraining stage. Subsequently, a boosting stage exploits this local information

to improve performance. The authors point out that training based on discriminative local appearances are more accurate compared to the usage of global image context. CNNs have also been proposed for the segmentation of isointense stage brain tissues [131] and brain extraction from multimodality MR images [56].

Hybrid approaches that combine CNNs with other architectures are also proposed. In [66], a deep learning algorithm is employed to encode the parameters of a deformable model and thus facilitate the segmentation of the left ventricle (LV) from short-axis cardiac MRI. CNNs are employed to automatically detect the LV, whereas deep Autoencoders are utilized to infer its shape. Yu *et al.* [67] designed a wireless capsule endoscopy classification system based on a hybrid CNN with extreme learning machine (ELM). The CNN constitutes a data-driven feature extractor, whereas the cascaded ELM acts as a strong classifier.

A comparison between different CNNs architectures concluded that deep CNNs of up to 22 layers can be useful even with limited training datasets [73]. More detailed description of various CNNs architectures proposed in medical imaging analysis is presented in previous survey [58]. The key challenges and limitations are:

1) CNNs are designed for 2-D images whereas segmentation problems in MRI and CT are inherently 3-D. This problem is further complicated by the anisotropic voxel size. Although the creation of isotropic images by interpolating the data is a possibility, it can result in severely blurred images. Another solution is to train the CNNs on orthogonal patches extracted from axial, sagittal and coronal views [62], [132]. This approach also drastically reduces the time complexity required to process 3-D information and thus alleviates the problem of overfitting.

2) CNNs do not model spatial dependencies. Therefore, several approaches have incorporated voxel neighboring information either implicitly or by adding a pairwise term in the cost function, which is referred as conditional random field [85].

3) Preprocessing to bring all subjects and imaging modalities to similar distribution is still a crucial step that affects the classification performance. Similarly to conventional machine learning approaches, balancing the datasets with bootstrapping and selecting samples with high entropy is advantageous.

Perhaps, all of these limitations result from or are exacerbated by small and incomplete training datasets. Furthermore, there is limited availability of ground-truth/annotated data, since the cost and time to collect and manually annotate medical images is prohibitively large. Manual annotations are subjective and highly variable across medical experts. Although, it is thought that the manual annotation would require highly specialized knowledge in medicine and medical imaging physics, recent studies suggest that nonprofessional users could perform similarly [76]. Therefore, crowdsourcing is suggested as a viable alternative to create low-cost, big ground-truth medical imaging datasets. Moreover, the normal class is often over represented since the healthy tissue usually dominates and forms highly repetitive pat-

terns. These issues result in slow convergence and overfitting. To alleviate the lack of training samples, transfer learning via fine tuning have been suggested in medical imaging applications [58], [72], [74], [76]. In transfer learning via fine-tuning, a CNN is pretrained using a database of labeled natural images. The use of natural images to train CNNs in medical imaging is controversial because of the profound difference between natural and medical images. Nevertheless, Tajbakhsh *et al.* [74] showed that fine-tuned CNNs based on natural images are less prone to overfitting due to the limited size training medical imaging sets and perform similarly or better than CNNs trained from scratch. Shin *et al.* [73] has applied transfer learning from natural images in thoraco-abdominal lymph node detection and interstitial lung disease classification. They also reported better results than training the CNNs from scratch with more consistent performances of validation loss and accuracy traces. Chen *et al.* [72] applied successfully a transfer learning strategy to identify the fetal abdominal standard plane. The lower layers of a CNN are pretrained based on natural images. The approach shows improved capability of the algorithm to encode the complicated appearance of the abdominal plane. Multitask training has also been suggested to handle the class imbalance common in CAD applications. Multitasking refers to the idea of solving different classification problems simultaneously and it results in a drastic reduction of free parameters [133].

Although CNNs have dominated medical image analysis applications, other deep learning approaches/architectures have also been applied successfully. In a recent paper, a stacked denoising autoencoder was proposed for the diagnosis of benign malignant breast lesions in ultrasound images and pulmonary nodules in CT scans [77]. The method outperforms classical CAD approaches, largely due to the automatic feature extraction and noise tolerance. Furthermore, it eliminates the image segmentation process to obtain a lesion boundary. Shan *et al.* [53] presented a stacked sparse autoencoder for microaneurysms detection in fundus images as an instance of a diabetic retinopathy strategy. The proposed method learns high-level distinguishing features based only on pixel intensities.

Various autoencoder-based learning approaches have also been applied to the automatic extraction of biomarkers from brain images and the diagnosis of neurological diseases. These methods often use available public domain brain image databases such as the Alzheimer's disease neuroimaging initiative database. For example, a deep Autoencoder combined with a softmax output layer for regression is proposed for the diagnosis of Alzheimer's disease. Hu *et al.* [134] also used autoencoders for Alzheimer's disease prediction based on Functional Magnetic Resonance Images (fMRI). The results show that the proposed method achieves much better classification than the traditional means. On the other hand, Li *et al.* [61] proposed an RBM approach that identifies biomarkers from MRI and positron emission tomography (PET) scans. They obtained an improvement of about 6% in classification accuracy compared to the standard approaches. Kuang *et al.* [60] proposed an RBM approach for fMRI data to discriminate attention deficit hyperactivity disorder. The system is capable of predicting the subjects as control, combined, inattentive or hyperactive through their

frequency features. Suk *et al.* [59] proposed a DBM to extract a latent hierarchical feature representation from 3-D patches of brain images.

Low level image processing, such as image segmentation and registration can also benefit from deep learning models. Brosch *et al.* [64] described a manifold learning approach of 3-D brain images based on DBN. It is different than other methods because it does not require a locally linear manifold space. Mansoor *et al.* [54] developed a fully automated shape model segmentation mechanism for the analysis of cranial nerve systems. The deep learning approach outperforms conventional methods particularly in regions with low contrast, such as optic tracts and areas with pathology. In [135], a pipeline is proposed for object detection and segmentation in the context of automatically processing volumetric images. A novel framework called marginal space deep learning implements an object parameterization in hierarchical marginal spaces combined with automatic feature detection based on deep learning. In [84], a DNN architecture called input–output deep architecture is described to solve the image labelling problem. A single NN forward step is used to assign a label to each pixel. This method avoids the hand-crafted subjective design of a model with a deep learning mechanism, which automatically extracts the dependencies between labels. Deep learning is also used for processing hyperspectral images [83]. Spectral and spatial learned features are combined together in a hierarchical model to characterize tissues or materials.

In [78], a hybrid multilayered group method of data handling, which is a special NN with polynomial activation functions, has been used together with a principal component-regression analysis to recognize the liver and spleen. A similar approach is used for the identification of the myocardium [79] as well as the right and left kidney regions [80]. The authors extend the method to analyze brain or lung CT images to detect cancer [81]. Zhen *et al.* [63] presents a framework for direct biventricular volume estimation, which avoids the need of user inputs and over simplification assumptions. The learning process involves unsupervised cardiac image representation with multiscale deep networks and direct biventricular volume estimation with RF. Rose *et al.* [82] propose a methodology for hierarchical clustering in application to mammographic image data. Classification is performed based on a deep learning architecture along with a standard NN.

In general, deep learning in medical imaging provides automatic discovery of object features and automatic exploration of feature hierarch and interaction. In this way, a relatively simple training process and a systematic performance tuning can be used, making deep learning approaches improve over the state-of-the art. However, in medical imaging analysis, their potentials have not been unfolded fully. To be successful in disease detection and classification approaches, deep learning requires the availability of large labeled datasets. Annotating imaging datasets is an extremely time-consuming and costly process that is normally undertaken by medical doctors. Currently, there is a lot of debate on whether to increase the number of annotated datasets with the help of non-experts (crowd-sourcing) and how to standardize the available images to allow objective assessment of the deep learning approaches.
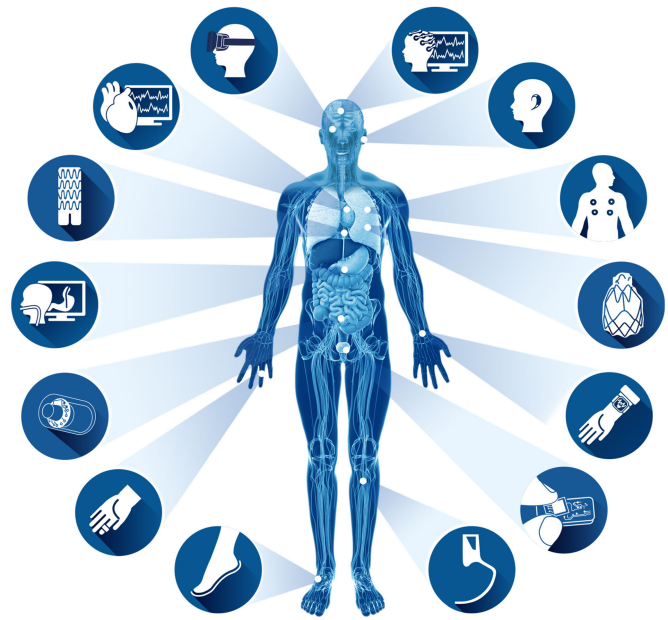

Fig. 6. Data for health monitoring applications can be captured using a wide array of pervasive sensors that are worn on the body, implanted, or captured through ambient sensors, e.g., inertial motion sensors, ECG patches, smart-watches, EEG, and prosthetics.

C. Pervasive Sensing for Health and Wellbeing

Pervasive sensors, such as wearable, implantable, and ambient sensors [136] allow continuous monitoring of health and wellbeing, Fig. 6. An accurate estimation of food intake and energy expenditure throughout the day, for example, can help tackle obesity and improve personal wellbeing. For elderly patients with chronic diseases, wearable and ambient sensors can be utilized to improve quality of care by enabling patients to continue living independently in their own homes. The care of patients with disabilities and patients undergoing rehabilitation can also be improved through the use of wearable and implantable assistive devices and human activity recognition. For patients in critical care, continuous monitoring of vital signs, such as blood pressure, respiration rate, and body temperature, are important for improving treatment outcomes by closely analyzing the patient's condition [137].

1) Energy Expenditure and Activity Recognition: Obesity has been identified as an escalating global epidemic health problem and is found to be associated with many chronic diseases, including type 2 diabetes and cardiovascular diseases. Dietitian recommend that only a standard amount of calories should be consumed to maintain a healthy balance within the body. Accurately recording the foods consumed and physical activities performed can help to improve health and manage diseases; however, selecting features that can generalize across the wide variety of food and daily activities is a major challenge. A number of solutions that use smartphones or wearable devices have been proposed for managing food intake and monitoring energy expenditure.

In [99], an assistive calorie measurement system is proposed to help patients and doctors to control diet-related health conditions. The proposed smartphone-based system estimates the calories contained in pictures of food taken by the user. In

order to recognize food accurately in the system, a CNN is used. In [100], deep learning, mobile cloud computing, distance estimation, and size calibration tools are implemented on a mobile device for food calorie estimation.

To identify different activities, [90] proposes to combine deep learning techniques with invariant and slowly varying features for the purpose of learning hierarchical representations from video. Specifically, it uses a two-layered structure with 3-D convolution and max pooling to make the method scalable to large inputs. In [94], a deep learning based algorithm is developed for human activity recognition using RGB-D video sequences. A temporal structure is learnt in order to improve the classification of human activities. [91] proposed an elderly and child care intelligent surveillance system where a three stream CNN is proposed for recognizing particular human actions such as fall and baby crawl. If the system detects abnormal activities, it will raise an alarm and notify family members.

Zeng *et al.* [92] compared the performance of a CNN based method on three public human activity recognition datasets and found that their deep learning approach can obtain better overall classification accuracy across different human activities as the method is more generalizable. Ha *et al.* [93] also used a CNN for human activity recognition. CNNs can capture local relationships from data as well as provide invariance against distortion, which makes it popular for learning features from images and speech. Choi *et al.* [95] employed RBMs to learn activities using data from smart watches and home activity datasets, respectively, with improvements shown over baseline methods. However, for low-power devices such as smart-watches and sensor nodes, efficiency is often a concern, especially when a deep learning method with high computational complexity is needed for learning. To overcome this, Ravì *et al.* [96] proposed data preprocessing techniques to standardize and reduce variations in the input data caused by differences in sensor properties, such as placement and orientation.

2) Assistive Devices: Recognizing generic objects from the 3-D world, understanding shape and volume or classification of scene are important features required for assistive devices. These applications are mainly developed to guide users and provide audio or tactile feedback, for example, in the case of impaired patients that need a system to avoid obstacles along the path or receive information concerned with the surrounding environment. For example, Poggi *et al.* [97] proposed a robust obstacle detection system for people suffering from visual impairments. Here a wearable device based on CNN is designed.

Assistive devices that can recognize hand gestures have also been proposed for patients with disabilities—for applications such as sign language interpretation—and sterile environments in the surgical setting—to allow for touch less human-computer-interaction (HRI). However, gesture recognition is a very challenging task due to the complexity and large variations in hand postures. Huang *et al.* [98] proposes a method for sign language recognition which involves the use of a DNN fed with real-sense data. The DNN takes the 3-D coordinates of finger joints as inputs directly with no handcrafted features used.

3) Detection of Abnormalities in Vital Signs: For critically ill patients, identifying abnormalities in their vital signs

is important. These episodes, however, are rare, vary between patients, and susceptible to noise and artifacts. Machine learning approaches have been proposed for detecting abnormalities under a varying set of condition and thus their application in a clinical setting is limited. Furthermore, with continuous sensing, large volumes of data can be generated, such as electroencephalography (EEG) record signal from a large number of input channels with a high temporal resolution (several kHz). Managing this amount of time-series data requires the development of online algorithms that could process the varying types of data.

Wulsin *et al.* [89] proposed a DBN approach to detect anomalies in EEG waveforms. EEG is used to record electrical activity of the brain. Interpreting the waveforms from brain activity is challenging due to the high dimensionality of the input signal and the limited understanding of the intrinsic brain operations. Using a large set of training data, DBNs outperform SVM and have a faster query time of around 10s for 50 000 samples. Jia *et al.* [86] used a deep learning method based on RBMs to recognize affective state of EEG. Although the sample sets are small and noisy, the proposed method achieves greater accuracy. A DBN was also used for detecting arrhythmias from electrocardiography (ECG) signals. A DBN was also used in monitoring heart rhythm based on ECG data [87]. The main purpose of the system is identifying arrhythmias which are a complex pattern recognition problem. Yan *et al.* attained classification accuracies of 98% using a two-lead ECG dataset. For low-power wearable and implantable EEG sensors, where energy consumption and efficiency are major concerns, Wang *et al.* [88] designed a DBN to compress the signal. This results in more than 50% of energy savings while retaining accuracy for neural decoding.

The introduction of deep learning has increased the utility of pervasive sensing across a range of health applications by improving the accuracy of sensors that measure food calorie intake, energy expenditure, activity recognition, sign language interpretation, and detection of anomalous events in vital signs. Many applications use deep learning to achieve greater efficiency and performance for real-time processing on low-power devices; however, a greater focus should be placed upon implementations on neuromorphic hardware platforms designed for low-power parallel processing. The most significant improvements in performance have been achieved where the data has high dimensionality—as seen in the EEG datasets—or high variability—due to changes in sensor placement, activity, and subject. Most current research has focused on the recognition of activities of daily living and brain activity. Many opportunities for other applications and diseases remain, and many currently studies still rely upon relatively small datasets that may not fully capture the variability of the real world.

D. Medical Informatics

Medical Informatics focuses on the analysis of large, aggregated data in health-care settings with the aim to enhance and develop clinical decision support systems or assess medical data both for quality assurance and accessibility of health care services. Electronic health records (EHR) are an extremely rich

source of patient information, which include medical history details such as diagnoses, diagnostic exams, medications and treatment plans, immunization records, allergies, radiology images, sensors multivariate times series (such as EEG from intensive care units), laboratory, and test results. Efficient mining of this big data would provide valuable insight into disease management [138], [139]. Nevertheless, this is not trivial because of several reasons:

1) Data complexity owing to varying length, irregular sampling, lack of structured reporting and missing data. The quality of reporting varies considerably among institutions and persons.
2) Multimodal datasets of several petabytes that includes medical images, sensors data, lab results, and unstructured text reports.
3) Long-term time dependencies between clinical events and disease diagnosis and treatment that complicates learning. For example, long and varying delays often separate the onset of disease from the appearance of symptoms.
4) Inability of traditional machine learning approaches to scale up to large and unstructured datasets.
5) Lack of interpretability of results hinders adaptation of the methods in the clinical setting.

Deep learning approaches have been designed to scale up well with big and distributed datasets. The success of DNNs is largely due to their ability to learn novel features/patterns and understand data representation in both an unsupervised and supervised hierarchical manners. DNNs have also proven to be efficient in handling multimodal information since they can combine several DNN architectural components. Therefore, it is unsurprising that deep learning has quickly been adopted in medical informatics research. For example, Shin *et al.* [105] presented a combined text-image CNN to identify semantic information that links radiology images and reports from a typical picture archiving and communication system hospital system. Liang *et al.* [107] used a modified version of CDBN as an effective training method for large-scale datasets on hypertension, and Chinese medical diagnosis from a manually converted EHR database. Putin *et al.* [108] applied DNNs for identifying markers that predict human chronological age based on simple blood tests. Nie *et al.* [103] proposed a deep learning network for automatic disease inference, which requires manual gathering the key symptoms or questions related to the disease.

In another study, Mioto *et al.* [102] showed that a stack of denoising autoencoders can be used to automatically infer features from a large-scale EHR database and represent patients without requiring additional human effort. These general features can be used in several scenarios. The authors demonstrated the ability of their system to predict the probability of a patient developing specific diseases, such as diabetes, schizophrenia and cancer. Furthermore, Futoma *et al.* [109] compared different models in their ability to predict hospital readmissions based on a large EHR database. DNNs have significantly higher prediction accuracies than conventional approaches, such as penalized logistic regression, though training of the DNN models were not straightforward.

To tackle time dependencies in EHR with multivariate time series from intensive care monitoring systems, Lipton *et al.* [106] employed a LSTM RNN. The reason for using RNNs is that their ability to memorize sequential events could improve the modeling of the varying time delays between the onsets of emergency clinical events, such as respiratory distress and asthma attack and the appearance of symptoms. In a related study, Mehrabi *et al.* [104] proposed the use DBN to discover common temporal patterns and characterize disease progression. The authors highlighted that the ability to discern and interpret the newly discovered patterns requires further investigation.

The motivations behind these studies are to develop general purpose systems to accurately predict length of stay, future illness, readmission, and mortality with the view to improve clinical decision making and optimize clinical pathways. Early prediction in health care is directly related to saving patients' lives. Furthermore, the discovery of novel patterns can result in new hypotheses and research questions. In computational phenotyping research, the goal is to discover meaningful data-driven features and disease characteristics.

For example, Che *et al.* [101] highlighted that although DNNs outperform conventional machine learning approaches in their ability to predict and classify clinical events, they suffer from the issue of model interpretability, which is important for clinical adaptation. They pointed out that interpreting individual units can be misleading and the behavior of DNNs are more complex than originally thought. They suggested that once a DNN is trained with big data, a simpler model can be used to distil knowledge and mimic the prediction performance of the DNN. To interpret features from deep learning models such as stacked denoising autoencoder and LSTM RNNs, they use gradient boosting decision trees (GBDT). GBDT are an ensemble of weak prediction models and in this work they represent a linear combination of functions.

Deep learning has paved the way for personalized health care by offering an unprecedented power and efficiency in mining large multimodal unstructured information stored in hospitals, cloud providers and research organization. Although, it has the potential to outperform traditional machine learning approaches, appropriate initialization and tuning is important to avoid overfitting. Noisy and sparse datasets result in considerable fall of performance indicating that there are several challenges to be addressed. Furthermore, adopting these systems into clinical practice requires the ability to track and interpret the extracted features and patterns.

E. Public Health

Public health aims to prevent disease, prolong life, and promote healthcare by analyzing the spread of disease and social behaviors in relation to environmental factors. Public health studies consider small localized populations to large populations that encompass several continents such as in the case of epidemics and pandemics. Applications involve epidemic surveillance, modeling lifestyle diseases, such as obesity, with relation to geographical areas, monitoring and predicting air quality, drug safety surveillance, etc. The conventional predic-

tive models scale exponentially with the size of the data and use complex models derived from physics, chemistry, and biology. Therefore, tuning these systems depend on parameterizations and ad hoc twists that only experts can provide. Nevertheless, existing computational methods are able to accurately model several phenomena, including the progression of diseases or the spread of air pollution. However, they have limited abilities in incorporating real time information, which could be crucial in controlling an epidemic or the adverse effects of a newly approved medicine. In contrast, deep learning approaches have a powerful generalization ability. They are data-driven methods that automatically build a hierarchical model and encode the information within their structure. Most deep learning algorithm designs are based on online machine learning and, thus, optimization of the cost function takes place sequentially as new training datasets become available. One of the simplest online optimization algorithms applied in DNNs is stochastic gradient descent. For these reasons, deep learning, along with recommendation systems and network analysis, are suggested as the key analysis methods for public health studies [140].

For example, monitoring and forecasting the concentration of air pollutants represents an area where deep learning has been successful. Ong *et al.* [110] reports that poor air quality is responsible for around 60 000 annual deaths and it is the leading cause for a number of chronic obstructive pulmonary diseases. They describe a system to predict the concentration of major air pollutant substances in Japan based on sensor data captured from over 52 cities. The proposed DNN consists of stacked Autoencoders and is trained in an online fashion. This deep architecture differs from the standard deep Autoencoders in that the output components are added gradually during training. To allow tracking of the large number of sensors and interpret the results, the authors exploited the sparsity in the data and they fine-tuned the DNN based on regularization approaches. Nevertheless, the authors pointed out that deep learning approaches as data-driven methods are affected by the inaccuracies and incompleteness of real-world data.

Another interesting application is tracking outbreaks with social media for epidemiology and lifestyle diseases. Social media can provide rich information about the progression of diseases, such as Influenza and Ebola, in real time. Zhao *et al.* [116] used the microblogging social media service, Twitter, to continuously track health states from the public. DNN is used to mine epidemic features that are then combined into a simulated environment to model the progression of disease. Text from Twitter messages can also be used to gain insight into antibiotics and infectious intestinal diseases. In [112], DBN is used to categorize antibiotic-related Twitter posts into nine classes (side effects, wanting/needing, advertisement, advice/information, animals, general use, resistance, misuse, and other). To obtain the classifier, Twitter messages were randomly selected for manual labeling and categorization. They used a training set of 412 manually labeled and 150 000 unlabeled examples. A deep learning approach based on RBMs was pretrained in a layer-by-layer procedure. Fine-tuning was based on standard back propagation and the labeled data. In [114], deep learning is used to create a topical vocabulary of keywords related to three types

of infectious intestinal disease—campylobacter, norovirus, and food poisoning. When compared to officially documented cases, their results show that social media can be a good predictor of intestinal diseases.

For tracking certain stigmatized behaviors, social media can also provide information that is often undocumented; Garimella *et al.* [115] used geographically-tagged images from Instagram to track lifestyle diseases, such as obesity, drinking, and smoking, and compare the self-categorization of images from the user against annotations obtained using a deep learning algorithm. The study found that while self-annotation generally provides useful demographic information, machine generated annotations were more useful for behaviors such as excessive drinking and substance abuse. In [111], a deep learning approach based on RBMs is designed to model and predict activity level and prevent obesity by taking into account self-motivation, social influences and environment events.

There is a growing interest in using mobile phone metadata to characterize and track human behavior. Metadata normally includes the duration and the location of the phone call or text message and it can provide valuable demographic information. A CNN was applied in predicting demographic information from mobile phone metadata, which was represented as temporal 2-D matrices. The CNN is comprised of a series of five horizontal convolution layers followed by a vertical convolution filter and two dense layers. The method provides high accuracy for age and gender prediction, whereas it eliminates the need for handcrafted features [113].

Mining the online data and metadata about individuals and large-scale populations via EHRs, mobile networks and social media is a means to inform public health and policy. Furthermore, mining food and drug records to identify adverse events could provide vital large scale alert mechanisms. We have presented a few examples that use deep learning for early identification and modeling the spread of epidemics and public health risks. However, strict regulation that protects data privacy limits the access and aggregation of the relevant information. For example, Twitter messages or Facebook posts could be used to identify new mothers at risk from postpartum depression. Although, this is positive, there is controversy associated of whether this information should become available, since it stigmatizes specific individuals. Therefore, it has become evident that we need to strike a balance between ensuring individuals can control access to their private medical information and providing pathways on how to make information available for public health studies [117]. The complexity and limited interpretability of deep learning models constitute an obstacle in allowing an informed decision about the precise operation of a DNN, which may limit its application in sensitive data.

IV. Deep Learning in Healthcare: Limitations and Challenges

Although for different artificial intelligence tasks, deep learning techniques can deliver substantial improvements in comparison to traditional machine learning approaches, many researchers and scientists remain sceptical of their use where

medical applications are involved. These scepticisms arise since deep learning theories have not yet provided complete solutions and many questions remain unanswered. The following four aspects summarize some of the potential issues associated with deep learning:

1) Despite some recent work on visualizing high level features by using the weight filters in a CNN [141], [142], the entire deep learning model is often not interpretable. Consequently, most researchers use deep learning approaches as a black box without the possibility to explain why it provides good results or without the ability to apply modifications in the case of misclassification issues.

2) As we have already highlighted in the previous sections, to train a reliable and effective model, large sets of training data are required for the expression of new concepts. Although recently we have witnessed an explosion of available healthcare data with many organizations starting to effectively transform medical records from paper to electronic records, disease specific data is often limited. Therefore, not all applications—particularly rare diseases or events—are well suited to deep learning. A common problem that can arise during the training of a DNN (especially in the case of small datasets) is overfitting, which may occur when the number of parameters in the network is proportional to the total number of samples in the training set. In this case, the network is able to memorize the training examples, but cannot generalize to new samples that it has not already observed. Therefore, although the error on the training set is driven to a very small value, the errors for new data will be high. To avoid the overfitting problem and improve generalization, regularization methods, such as the dropout [143], are usually exploited during training.

3) Another important aspect to take into account when deep learning tools are employed, is that for many applications the raw data cannot be directly used as input for the DNN. Thus, preprocessing, normalization or change of input domain is often required before the training. Moreover, the setup of many hyperparameters that control the architecture of a DNN, such as the size and the number of filter in a CNN, or its depth, is still a blind exploration process that usually requires accurate validation. Finding the correct preprocessing of the data and the optimal set of hyperparameters can be challenging, since it makes the training process even longer, requiring significant training resources and human expertise, without which is not possible to obtain an effective classification model.

4) The last aspect that we would like to underline is that many DNNs can be easily fooled. For example, [144] shows that it is possible to add small changes to the input samples (such as imperceptible noise in an image) to cause samples to be misclassified. However, it is important to note that almost all machine learning algorithms are susceptible to such issues. Values of particular features can be deliberately set very high or very low to induce misclassification in logistic regression. Simi-

larly, for decision tress, a single binary feature can be used to direct a sample along the wrong partition by simply switching it at the final layer. Hence in general, any machine learning models are susceptible to such manipulations. On the other hand, the work in [145] discusses the opposite problem. The author shows that it is possible to obtain meaningless synthetic samples that are strongly classified into classes even though they should not have been classified. This is also a genuine limitation of the deep learning paradigm, but it is a drawback for other machine learning algorithms as well.

To conclude, we believe that healthcare informatics today is a human-machine collaboration that may ultimately become a symbiosis in the future. As more data becomes available, deep learning systems can evolve and deliver where human interpretation is difficult. This can make diagnoses of diseases faster and smarter and reduce uncertainty in the decision making process. Finally, the last boundary of deep learning could be the feasibility of integrating data across disciplines of health informatics to support the future of precision medicine.

V. CONCLUSION

Deep learning has gained a central position in recent years in machine learning and pattern recognition. In this paper, we have outlined how deep learning has enabled the development of more data-driven solutions in health informatics by allowing automatic generation of features that reduce the amount of human intervention in this process. This is advantageous for many problems in health informatics and has eventually supported a great leap forward for unstructured data such as those arising from medical imaging, medical informatics, and bioinformatics. Until now, most applications of deep learning to health informatics have involved processing health data as an unstructured source. Nonetheless, a significant amount of information is equally encoded in structured data such as EHRs, which provide a detailed picture of the patient's history, pathology, treatment, diagnosis, outcome, and the like. In the case of medical imaging, the cytological notes of a tumor diagnosis may include compelling information like its stage and spread. This information is beneficial to acquire a holistic view of a patient condition or disease and then be able to improve the quality of the obtained inference. In fact, robust inference through deep learning combined with artificial intelligence could ameliorate the reliability of clinical decision support systems. However, several technical challenges remain to be solved. Patient and clinical data is costly to obtain and healthy control individuals represent a large fraction of a standard health dataset. Deep learning algorithms have mostly been employed in applications where the datasets were balanced, or, as a work-around, in which synthetic data was added to achieve equity. The later solution entails a further issue as regards the reliance of the fabricated biological data samples. Therefore, methodological aspects of NNs need to be revisited in this regard. Another concern is that deep learning predominantly depends on large amounts of training data. Such requirements make more critical the classical entry barriers of machine learning, i.e., data availability and privacy. Consequently, advances

in the development of seamless and fast equipment for health monitoring and diagnoses will play a prominent role in future research. Reference to the issue of computational power, we envisage that for the years to come, further ad hoc hardware platforms for neural networks and deep learning processing will be announced and made commercially available. It is worth noting that the rise of deep learning has been mightily supported by major IT companies (e.g., Google, Facebook, and Baidu) which hold a large extent of patents in the field and core businesses are substantially supported by data gathering, enormous storehouses and processing machines. Many researchers have been encouraged to apply deep learning to any data-mining and pattern recognition problem related to health informatics in light of the wide availability of free packages to support this research. Looking at it from the bright side, it has fostered an interesting trend and boosted the expectations of what machine learning could achieve on its own. Nevertheless, we should not consider deep learning as a silver bullet for every single challenge set by health informatics. In practice, it is still questionable whether the large amount of training data and computational resources needed to run deep learning at full performance is worthwhile, considering other fast learning algorithms that may produce close performance with fewer resources, less parameterization, tuning, and higher interpretability. Therefore, we conclude that deep learning has provided a positive revival of NNs and connectionism from the genuine integration of the latest advances in parallel processing enabled by coprocessors. Nevertheless, a sustained concentration of health informatics research exclusively around deep learning could slow down the development of new machine learning algorithms with a more conscious use of computational resources and interpretability.

REFERENCES

[1] H. R. Roth *et al.*, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1170–1181, May 2016.

[2] R. Fakoor, F. Ladhak, A. Nazi, and M. Huber, "Using deep learning to enhance cancer diagnosis and classification," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1–7.

[3] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, "Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning," *Nature Biotechnol.*, vol. 33, pp. 831–838, 2015.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[5] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[6] C. Poultney *et al.*, "Efficient learning of sparse representations with an energy-based model," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 1137–1144.

[7] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.

[8] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, "Contractive auto-encoders: Explicit invariance during feature extraction," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 833–840.

[9] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proc. Int. Conf. Artif. Neural Netw.*, 2011, pp. 52–59.

[10] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[11] R. Salakhutdinov and G. E. Hinton, "Deep boltzmann machines." in *Proc. Int. Conf. Artif. Intell. Stat.*, 2009, vol. 1, Art. no. 3.

[12] L. Younes, "On the convergence of markovian stochastic algorithms with rapidly decreasing ergodicity rates," *Stochastics: An Int. J. Probab. Stochastic Process.*, vol. 65, no. 3/4, pp. 177–228, 1999.

[13] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Comput.*, vol. 1, no. 2, pp. 270–280, 1989.

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[15] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, 1962.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[17] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vision*, 2014, pp. 818–833.

[18] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[19] R. Frank, "The perceptron a perceiving and recognizing automaton," Cornell Aeronautical Laboratory, Buffalo, NY, USA, Tech. Rep. 85-460-1, 1957.

[20] J. L. McClelland *et al.*, *Parallel distributed processing*. Cambridge, MA, USA: MIT Press, vol. 2, 1987.

[21] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," in *Neurocomputing: Foundations of Research*, J. A. Anderson and E. Rosenfeld, Eds. Cambridge, MA, USA: MIT Press, 1988, pp. 696–699. [Online]. Available: http://dl.acm.org/citation.cfm?id=65669.104451

[22] J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, Q. V. Le, and A. Y. Ng, "On optimization methods for deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 265–272.

[23] P. Domingos, "A few useful things to know about machine learning," *Commun. ACM*, vol. 55, no. 10, pp. 78–87, 2012.

[24] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–999, Sep. 1999.

[25] C. M. Bishop, "Pattern recognition," *Mach. Learn.*, vol. 128, pp. 1–737, 2006.

[26] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[28] Center Berkeley, "Caffe," 2016. [Online]. Available: http://caffe.berkeley vision.org/

[29] Microsoft, "Cntk," 2016. [Online]. Available: https://github.com/Microsoft/CNTK.

[30] Skymind, "Deeplearning4j," 2016.[Online]. Available: http://deeplearning4j.org/

[31] Wolfram Research, "Wolfram math," 2016. [Online]. Available: https://www.wolfram.com/mathematica/

[32] Google, "Tensorflow," 2016. [Online]. Available: https://www.tensorflow.org/

[33] Universite de Montreal, "Theano," 2016. [Online]. Available: http://deeplearning.net/software/theano/

[34] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch," 2016. [Online]. Available: http://http://torch.ch/

[35] Franois Chollet, "Keras," 2016. [Online]. Available: https://keras.io/

[36] Nervana Systems, "Neon," 2016. [Online]. Available: https://github.com/NervanaSystems/neon

[37] D. Ackely, G. Hinton, and T. Sejnowski, "Learning and relearning in boltzman machines," in *Parallel Distributed Processing: Explorations in Microstructure of Cognition*. Cambridge, MA, USA: MIT Press, pp. 282–317, 1986.

[38] H. Wang and D.-Y. Yeung, "Towards Bayesian deep learning: A survey," ArXiv e-prints, Apr. 2016.

[39] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Mateo, CA, USA: Morgan Kaufmann, 2014.

[40] M. A. Carreira-Perpinan and G. Hinton, "On contrastive divergence learning." in *Proc. Int. Conf. Artif. Intell. Stat.*, 2005, vol. 10, pp. 33–40.

[41] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomput.*, vol. 187, pp. 27–48, 2016.

[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[43] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proc. Int. Conf. Mach. Learn.*, 2009, pp. 609–616.

[44] NVIDIA corp., "Nvidia dgx-1," 2016. [Online]. Available: http://www.nvidia.com/object/deep-learning-system.html

[45] L. A. Pastur-Romay, F. Cedrón, A. Pazos, and A. B. Porto-Pazos, "Deep artificial neural networks and neuromorphic chips for big data analysis: Pharmaceutical and bioinformatics applications," *Int. J. Molecular Sci.*, vol. 17, no. 8, 2016, Art. no. 1313.

[46] R. Ibrahim, N. A. Yousri, M. A. Ismail, and N. M. El-Makky, "Multi-level gene/mirna feature selection using deep belief nets and active learning," in *Proc. Eng. Med. Biol. Soc.*, 2014, pp. 3957–3960.

[47] M. Khademi and N. S. Nedialkov, "Probabilistic graphical models and deep belief networks for prognosis of breast cancer," in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl.*, 2015, pp. 727–732.

[48] D. Quang, Y. Chen, and X. Xie, "Dann: A deep learning approach for annotating the pathogenicity of genetic variants," *Bioinformatics*, vol. 31, p. 761–763, 2014.

[49] B. Ramsundar, S. Kearnes, P. Riley, D. Webster, D. Konerding, and V. Pande, "Massively multitask networks for drug discovery," ArXiv e-prints, Feb. 2015.

[50] S. Zhang *et al.*, "A deep learning framework for modeling structural features of rna-binding protein targets," *Nucleic Acids Res.*, vol. 44, no. 4, pp. e32–e32, 2016.

[51] K. Tian, M. Shao, S. Zhou, and J. Guan, "Boosting compound-protein interaction prediction by deep learning," in *Proc. IEEE Int. Conf. Bioinformat. Biomed.*, 2015, pp. 29–34.

[52] C. Angermueller, H. Lee, W. Reik, and O. Stegle, "Accurate prediction of single-cell dna methylation states using deep learning," *bioRxiv*, 2016, Art. no. 055715.

[53] J. Shan and L. Li, "A deep learning method for microaneurysm detection in fundus images," in *Proc. IEEE Connected Health, Appl., Syst. Eng. Technol.*, 2016, pp. 357–358.

[54] A. Mansoor *et al.*, "Deep learning guided partitioned shape model for anterior visual pathway segmentation," *IEEE Trans. Med. Imag.*, vol. 35, no. 8, pp. 1856–1865, Aug. 2016.

[55] D. Nie, H. Zhang, E. Adeli, L. Liu, and D. Shen, "3d deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients," in *Proc. MICCAI*, 2016, pp. 212–220. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-46723-8_25

[56] J. Kleesiek *et al.*, "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping," *NeuroImage*, vol. 129, pp. 460–469, 2016.

[57] B. Jiang, X. Wang, J. Luo, X. Zhang, Y. Xiong, and H. Pang, "Convolutional neural networks in automatic recognition of trans-differentiated neural progenitor cells under bright-field microscopy," in *Proc. Instrum. Meas., Comput., Commun. Control*, 2015, pp. 122–126.

[58] M. Havaei, N. Guizard, H. Larochelle, and P. Jodoin, "Deep learning trends for focal brain pathology segmentation in MRI," *CoRR*, vol. abs/1607.05258, 2016. [Online]. Available: http://arxiv.org/abs/1607.05258

[59] H.-I. Suk *et al.*, "Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis," *NeuroImage*, vol. 101, pp. 569–582, 2014.

[60] D. Kuang and L. He, "Classification on ADHD with deep learning," in *Proc. Cloud Comput. Big Data*, Nov. 2014, pp. 27–32.

[61] F. Li, L. Tran, K. H. Thung, S. Ji, D. Shen, and J. Li, "A robust deep model for improved classification of ad/mci patients," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 5, pp. 1610–1616, Sep. 2015.

[62] K. Fritscher, P. Raudaschl, P. Zaffino, M. F. Spadea, G. C. Sharp, and R. Schubert, "Deep neural networks for fast segmentation of 3d medical images," in *Proc. MICCAI*, 2016, pp. 158–165. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-46723-8_19

[63] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation," *Med. Image Anal.*, vol. 30, pp. 120–129, 2016.

[64] T. Brosch *et al.*, "Manifold learning of brain mris by deep learning," in *Proc. MICCAI*, 2013, pp. 633–640.

[65] T. Xu, H. Zhang, X. Huang, S. Zhang, and D. N. Metaxas, "Multimodal deep learning for cervical dysplasia diagnosis," in *Proc. MICCAI*, 2016, pp. 115–123. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-46723-8_14

[66] M. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri," *Med. Image Anal.*, vol. 30, pp. 108–119, 2016.

[67] J.-s. Yu, J. Chen, Z. Xiang, and Y.-X. Zou, "A hybrid convolutional neural networks with extreme learning machine for WCE image classification," in *Proc. IEEE Robot. Biomimetics*, 2015, pp. 1822–1827.

[68] H. R. Roth *et al.*, "Anatomy-specific classification of medical images using deep convolutional nets," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2015, pp. 101–104.

[69] M. J. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1273–1284, May 2016.

[70] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1207–1216, May 2016.

[71] Y. Cao *et al.*, "Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor and marginalized communities," in *IEEE Connected Health, Appl., Syst. Eng. Technol.*, 2016, pp. 274–281.

[72] H. Chen *et al.*, "Standard plane localization in fetal ultrasound via domain transferred deep neural networks," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 5, pp. 1627–1636, Sep. 2015.

[73] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[74] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[75] Z. Yan *et al.*, "Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1332–1343, May 2016.

[76] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[77] J.-Z. Cheng *et al.*, "Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in ct scans," *Sci. Rep.*, vol. 6, 2016, Art. no. 24454.

[78] T. Kondo, J. Ueno, and S. Takao, "Medical image recognition of abdominal multi-organs by hybrid multi-layered GMDH-type neural network using principal component-regression analysis," in *Proc. 2nd Int. Symp. Comput. Netw.*, 2014, pp. 157–163.

[79] T. Kondo, U. Junji, and S. Takao, "Hybrid feedback GMDH-type neural network using principal component-regression analysis and its application to medical image recognition of heart regions," in *Proc. Joint 7th Int. Conf. Adv. Intell. Syst., 15th Int. Symp. Soft Comput. Intell. Syst.*, 2014, pp. 1203–1208.

[80] T. Kondo, S. Takao, and J. Ueno, "The 3-dimensional medical image recognition of right and left kidneys by deep GMDH-type neural network," in *Proc. Int. Conf. Intell. Informat. Biomed. Sci.*, 2015, pp. 313–320.

[81] T. Kondo, J. Ueno, and S. Takao, "Medical image diagnosis of lung cancer by deep feedback GMDH-type neural network," *Robot. Netw. Artif. Life*, vol. 2, no. 4, pp. 252–257, 2016.

[82] D. C. Rose, I. Arel, T. P. Karnowski, and V. C. Paquit, "Applying deep-layered clustering to mammography image analytics," in *Proc. Biomed. Sci. Eng. Conf.*, 2010, pp. 1–4.

[83] Y. Zhou and Y. Wei, "Learning hierarchical spectral-spatial features for hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 46, no. 7, pp. 1667–1678, Jul. 2016.

[84] J. Lerouge, R. Herault, C. Chatelain, F. Jardin, and R. Modzelewski, "Ioda: an input/output deep architecture for image labeling," *Pattern Recognit.*, vol. 48, no. 9, pp. 2847–2858, 2015.

[85] J. Wang, J. D. MacKenzie, R. Ramachandran, and D. Z. Chen, "A deep learning approach for semantic segmentation in histology tissue images," in *Proc. MICCAI*, 2016, pp. 176–184. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-46723-8_21

[86] X. Jia, K. Li, X. Li, and A. Zhang, "A novel semi-supervised deep learning framework for affective state recognition on eeg signals," in *Proc. Int. Conf. Bioinformat. Bioeng.*, 2014, pp. 30–37. [Online]. Available: http://dx.doi.org/10.1109/BIBE.2014.26

[87] Y. Yan, X. Qin, Y. Wu, N. Zhang, J. Fan, and L. Wang, "A restricted Boltzmann machine based two-lead electrocardiography classification," in *Proc. 12th Int. Conf. Wearable Implantable Body Sens. Netw.*, Jun. 2015, pp. 1–9.

[88] A. Wang, C. Song, X. Xu, F. Lin, Z. Jin, and W. Xu, "Selective and compressive sensing for energy-efficient implantable neural decoding," in *Proc. Biomed. Circuits Syst. Conf.*, Oct. 2015, pp. 1–4.

[89] D. Wulsin, J. Blanco, R. Mani, and B. Litt, "Semi-supervised anomaly detection for eeg waveforms using deep belief nets," in *Proc. 9th Int. Conf. Mach. Learn. Appl.*, Dec. 2010, pp. 436–441.

[90] L. Sun, K. Jia, T.-H. Chan, Y. Fang, G. Wang, and S. Yan, "DL-SFA: Deeply-learned slow feature analysis for action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2625–2632.

[91] C.-D. Huang, C.-Y. Wang, and J.-C. Wang, "Human action recognition system for elderly and children care using three stream convnet," in *Proc. Int. Conf. Orange Technol.*, 2015, pp. 5–9.

[92] M. Zeng *et al.*, "Convolutional neural networks for human activity recognition using mobile sensors," in *Proc. MobiCASE*, Nov. 2014, pp. 197–205. [Online]. Available: http://dx.doi.org/10.4108/icst. mobicase.2014.257786

[93] S. Ha, J. M. Yun, and S. Choi, "Multi-modal convolutional neural networks for activity recognition," in *Proc. Int. Conf. Syst., Man, Cybern.*, Oct. 2015, pp. 3017–3022.

[94] H. Yalçın, "Human activity recognition using deep belief networks," in *Proc. Signal Process. Commun. Appl. Conf.*, 2016, pp. 1649–1652.

[95] S. Choi, E. Kim, and S. Oh, "Human behavior prediction for smart homes using deep learning," in *Proc. IEEE RO-MAN*, Aug. 2013, pp. 173–179.

[96] D. Ravi, C. Wong, B. Lo, and G. Z. Yang, "Deep learning for human activity recognition: A resource efficient implementation on low-power devices," in *Proc. 13th Int. Conf. Wearable Implantable Body Sens. Netw.*, Jun. 2016, pp. 71–76.

[97] M. Poggi and S. Mattoccia, "A wearable mobility aid for the visually impaired based on embedded 3d vision and deep learning," in *Proc. IEEE Symp. Comput. Commun.*, 2016, pp. 208–213.

[98] J. Huang, W. Zhou, H. Li, and W. Li, "Sign language recognition using real-sense," in *Proc. IEEE ChinaSIP*, 2015, pp. 166–170.

[99] P. Pouladzadeh, P. Kuhad, S. V. B. Peddi, A. Yassine, and S. Shirmohammadi, "Food calorie measurement using deep learning neural network," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. Proc.*, 2016, pp. 1–6.

[100] P. Kuhad, A. Yassine, and S. Shimohammadi, "Using distance estimation and deep learning to simplify calibration in food calorie measurement," in *Proc. IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl.*, 2015, pp. 1–6.

[101] Z. Che, S. Purushotham, R. Khemani, and Y. Liu, "Distilling knowledge from deep networks with applications to healthcare domain," ArXiv e-prints, Dec. 2015.

[102] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep patient: An unsupervised representation to predict the future of patients from the electronic health records," *Sci. Rep.*, vol. 6, pp. 1–10, 2016.

[103] L. Nie, M. Wang, L. Zhang, S. Yan, B. Zhang, and T. S. Chua, "Disease inference from health-related questions via sparse deep learning," *IEEE Trans. Knowl. Data Eng*, vol. 27, no. 8, pp. 2107–2119, Aug. 2015.

[104] S. Mehrabi *et al.*, "Temporal pattern and association discovery of diagnosis codes using deep learning," in *Proc. Int. Conf. Healthcare Informat.*, Oct. 2015, pp. 408–416.

[105] H. Shin, L. Lu, L. Kim, A. Seff, J. Yao, and R. M. Summers, "Interleaved text/image deep mining on a large-scale radiology database for automated image interpretation," *CoRR*, vol. abs/1505.00670, 2015. [Online]. Available: http://arxiv.org/abs/1505.00670

[106] Z. C. Lipton, D. C. Kale, C. Elkan, and R. C. Wetzel, "Learning to diagnose with LSTM recurrent neural networks," *CoRR*, vol. abs/1511.03677, 2015. [Online]. Available: http://arxiv.org/abs/1511.03677

[107] Z. Liang, G. Zhang, J. X. Huang, and Q. V. Hu, "Deep learning for healthcare decision making with emrs," in *Proc. Int. Conf. Bioinformat. Biomed.*, Nov 2014, pp. 556–559.

[108] E. Putin *et al.*, "Deep biomarkers of human aging: Application of deep neural networks to biomarker development," *Aging*, vol. 8, no. 5, pp. 1–021, 2016.

[109] J. Futoma, J. Morris, and J. Lucas, "A comparison of models for predicting early hospital readmissions," *J. Biomed. Informat.*, vol. 56, pp. 229–238, 2015.

[110] B. T. Ong, K. Sugiura, and K. Zettsu, "Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting pm2. 5," *Neural Comput. Appl.*, vol. 27, pp. 1–14, 2015.

[111] N. Phan, D. Dou, B. Piniewski, and D. Kil, "Social restricted Boltzmann machine: Human behavior prediction in health social networks," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2015, pp. 424–431.

[112] R. L. Kendra, S. Karki, J. L. Eickholt, and L. Gandy, "Characterizing the discussion of antibiotics in the Twittersphere: What is the bigger picture?" *J. Med. Internet Res.*, vol. 17, no. 6, 2015, Art. no. e154.

[113] B. Felbo, P. Sundsøy, A. Pentland, S. Lehmann, and Y.-A. de Montjoye, "Using deep learning to predict demographics from mobile phone metadata," Feb. 2016. [Online]. Available: http://arxiv.org/abs/1511.06660

[114] B. Zou, V. Lampos, R. Gorton, and I. J. Cox, "On infectious intestinal disease surveillance using social media content," in *Proc. 6th Int. Conf. Digit. Health Conf.*, 2016, pp. 157–161.

[115] V. R. K. Garimella, A. Alfayad, and I. Weber, "Social media image analysis for public health," in *Proc. CHIConf. Human Factors Comput. Syst.*, 2016, pp. 5543–5547. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858234

[116] L. Zhao, J. Chen, F. Chen, W. Wang, C.-T. Lu, and N. Ramakrishnan, "Simnest: Social media nested epidemic simulation via online semisupervised deep learning," in *Proc. IEEE Int. Conf. Data Mining*, 2015, pp. 639–648.

[117] E. Horvitz and D. Mulligan, "Data, privacy, and the greater good," *Science*, vol. 349, no. 6245, pp. 253–255, 2015.

[118] J. C. Venter *et al.*, "The sequence of the human genome," *Science*, vol. 291, no. 5507, pp. 1304–1351, 2001.

[119] E. S. Lander *et al.*, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.

[120] L. Hood and S. H. Friend, "Predictive, personalized, preventive, participatory (p4) cancer medicine," *Nature Rev. Clin. Oncol.*, vol. 8, no. 3, pp. 184–187, 2011.

[121] M. K. Leung, A. Delong, B. Alipanahi, and B. J. Frey, "Machine learning in genomic medicine: A review of computational problems and data sets," *Proc. IEEE*, vol. 104, no. 1, pp. 176–197, Jan. 2016.

[122] C. Angermueller, T. Pärnamaa, L. Parts, and O. Stegle, "Deep learning for computational biology," *Molecular Syst. Biol.*, vol. 12, no. 7, 2016, Art. no. 878.

[123] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, "Molecular graph convolutions: moving beyond fingerprints," *J. Comput. Aided Mol. Des.*, vol. 30, no. 8, pp. 595–608, 2016. [Online]. Available: http://dx.doi.org/10.1007/s10822-016-9938-8

[124] E. Gawehn, J. A. Hiss, and G. Schneider, "Deep learning in drug discovery," *Molecular Informat.*, vol. 35, no. 1, pp. 3–14, 2016.

[125] H. Hampel, S. Lista, and Z. S. Khachaturian, "Development of biomarkers to chart all alzheimer?s disease stages: The royal road to cutting the therapeutic gordian knot," *Alzheimer's Dementia*, vol. 8, no. 4, pp. 312–336, 2012.

[126] V. Marx, "Biology: The big challenges of big data," *Nature*, vol. 498, no. 7453, pp. 255–260, 2013.

[127] S. Ekins, "The next era: Deep learning in pharmaceutical research," *Pharmaceutical Res.*, vol. 33, no. 11, pp. 2594–2603, 2016.

[128] D. de Ridder, J. de Ridder, and M. J. Reinders, "Pattern recognition in bioinformatics," *Briefings Bioinformat.*, vol. 14, no. 5, pp. 633–647, 2013.

[129] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural Networks: Tricks of the Trade*. New York, NY, USA: Springer, 2012, pp. 437–478.

[130] H. Y. Xiong *et al.*, "The human splicing code reveals new insights into the genetic determinants of disease," *Science*, vol. 347, no. 6218, 2015, Art. no. 1254806.

[131] W. Zhang *et al.*, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.

[132] Y. Zheng, D. Liu, B. Georgescu, H. Nguyen, and D. Comaniciu, "3d deep learning for efficient and robust landmark detection in volumetric data," in *Proc. MICCAI*, 2015, pp. 565–572.

[133] A. Jamaludin, T. Kadir, and A. Zisserman, "Spinenet: Automatically pinpointing classification evidence in spinal MRIS," in *Proc. MICCAI*, 2016, pp. 166–175. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-46723-8_20

[134] C. Hu, R. Ju, Y. Shen, P. Zhou, and Q. Li, "Clinical decision support for alzheimer's disease based on deep learning and brain network," in *Proc. Int. Conf. Commun.*, May 2016, pp. 1–6.

[135] F. C. Ghesu *et al.*, "Marginal space deep learning: Efficient architecture for volumetric image parsing," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1217–1228, May 2016.

[136] G.-Z. Yang, *Body Sensor Networks*, 2nd ed. New York, NY, USA: Springer, 2014.

[137] A. E. W. Johnson, M. M. Ghassemi, S. Nemati, K. E. Niehaus, D. A. Clifton, and G. D. Clifford, "Machine learning and decision support in critical care," *Proc. IEEE*, vol. 104, no. 2, pp. 444–466, Feb. 2016.

[138] J. Andreu-Perez, C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong, and G. Z. Yang, "Big data for health," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 4, pp. 1193–1208, Jul. 2015.

[139] G.-Z. Yang and D. R. Leff, "Big data for precision medicine," *Engineering*, vol. 1, no. 3, 2015, Art. no. 277. [Online]. Available: http://engineering.org.cn/EN/abstract/article_12197.shtml

[140] T. Huang, L. Lan, X. Fang, P. An, J. Min, and F. Wang, "Promises and challenges of big data computing in health sciences," *Big Data Res.*, vol. 2, no. 1, pp. 2–11, 2015.

[141] D. Erhan, Y. Bengio, A. Courville, and P. Vincent, "Visualizing higher-layer features of a deep network," Univ. Montreal, Montreal, QC, Canada, Tech. Rep. 1341, 2009.

[142] D. Erhan, A. Courville, and Y. Bengio, "Understanding representations learned in deep architectures," Department dInformatique et Recherche Operationnelle, University of Montreal, QC, Canada, Tech. Rep. 1355, 2010.

[143] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting." *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[144] A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 427–436.

[145] C. Szegedy *et al.*, "Intriguing properties of neural networks." *CoRR*, vol. abs/1312.6199, 2013. [Online]. Available: http://dblp.uni-trier.de/db/journals/corr/corr1312.html#SzegedyZSBEGF13

Authors' photographs and biographies not available at the time of publication.