

# Deep Learning in Bioinformatics

Seonwoo Min<sup>1</sup>, Byunghan Lee<sup>1</sup>, and Sungroh Yoon<sup>1,2\*</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Seoul National University, Seoul 151-744, Korea

<sup>2</sup>Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-747, Korea

## Abstract

As we are living in the era of big data, transforming biomedical big data into valuable knowledge has been one of the most important problems in bioinformatics. At the same time, deep learning has advanced rapidly since early 2000s and is recently showing a state-of-the-art performance in various fields. So naturally, applying deep learning in bioinformatics to gain insights from data is under the spotlight of both the academia and the industry. This article reviews some research of deep learning in bioinformatics. To provide a big picture, we categorized the research by both bioinformatics domains — omics, biomedical imaging, biomedical signal processing — and deep learning architectures — deep neural network, convolutional neural network, recurrent neural network, modified neural network — as well as present brief descriptions of each work. Additionally, we introduce a few issues of deep learning in bioinformatics such as problems of class imbalance data and suggest future research directions such as multimodal deep learning. We believe that this paper could provide valuable insights and be a starting point for researchers to apply deep learning in their bioinformatics studies.

---

\*Corresponding author. Mailing address: 301-908, Department of Electrical and Computer Engineering, Seoul National University, Seoul 151-744, Korea. E-mail: [sryoon@snu.ac.kr](mailto:sryoon@snu.ac.kr). Phone: +82-2-880-1401.

## **Keywords**

Deep learning, neural network, machine learning, bioinformatics, omics, biomedical imaging, biomedical signal processing

## **Key Points**

- ◆ As a great deal of biomedical data has been accumulated, various machine algorithms have been widely applied in bioinformatics to extract knowledge from big data.
- ◆ Deep learning, emerging on the basis of big data, the power of parallel and distributed computing, and sophisticated algorithms, is making major advances in many domains such as image recognition, speech recognition, and natural language processing.
- ◆ Naturally, many studies have been conducted to apply deep learning in bioinformatics, including the domains of omics, biomedical imaging, and biomedical signal processing.
- ◆ This paper reviews research of deep learning in bioinformatics categorized by both bioinformatics domains and deep learning architectures — deep neural network, convolutional neural network, recurrent neural network, modified neural network — and presents common issues and future research directions.
- ◆ As an overall review of existing works, we believe that this paper could provide valuable insights and be helpful for researchers to apply deep learning in their bioinformatics research.

## **Author Description**

**Seonwoo Min** is a M.S./Ph.D. candidate at Department of Electrical and Computer Engineering, Seoul National University, Korea. His research areas include high-performance bioinformatics, machine learning for biomedical big data, and deep learning.

**Byunghan Lee** is a Ph.D. candidate at Department of Electrical and Computer Engineering, Seoul National University, Korea. His research areas include high-performance bioinformatics, machine learning for biomedical big data, and data mining.

**Sungroh Yoon** is an associate professor at Department of Electrical and Computer Engineering, Seoul National University, Seoul, Korea. He received his Ph.D. and postdoctoral training from Stanford University, Stanford, USA. His research interests include RNA bioinformatics, machine learning in bioinformatics, and high-performance bioinformatics.

## Introduction

As we are living in the era of big data, transforming big data into valuable knowledge has become important more than ever [1]. Certainly, bioinformatics is no exception in such trends. Various forms of biomedical data including omics data, image, and signal have been significantly accumulated, and its great potential in biological and health-care research has caught the interests of the industry as well as the academia. For instance, IBM provided Watson for Oncology, a platform analyzing patients' medical information and assisting clinicians with treatment options [2, 3]. In addition, Google DeepMind, achieving a great success with AlphaGo in the game of GO, recently launched DeepMind Health to develop effective health-care technologies [4, 5].

To extract knowledge from huge data in bioinformatics, machine learning has been one of the most widely used methodologies. Machine learning algorithms use training data to uncover underlying patterns, build a model, and then make predictions on the new data based on the model. Some of the well-known algorithms — support vector machine, hidden Markov model, Bayesian networks, Gaussian networks — have been applied in genomics, proteomics, systems biology, and many other domains [6].

Conventional machine learning algorithms have limitations in processing the raw form of data, so researchers put a tremendous effort in transforming the raw form into suitable high-abstraction level features with considerable domain expertise [7]. On the other hand, deep learning, a new type of machine learning algorithm, has emerged recently on the basis of big data, the power of parallel and distributed computing, and sophisticated algorithms. Deep learning algorithms have overcome the former limitations and are making major advances in diverse fields such as image recognition, speech recognition, and natural language processing. Certainly, bioinformatics is no exception in deep learning applications. Several studies have been conducted to apply deep learning in bioinformatics as in Figure 1. We categorized the research by the form of input data into three domains: omics, biomedical imaging, and biomedical signal processing. Detailed lists of bioinformatics research topics where deep learning is applied and input data examples of each domain are shown in Table 1.

**Table 1:** Deep learning applied bioinformatics research topics and input data

	Input data	Research topics
<b>Omics</b>	sequencing data (DNA-seq, RNA-seq, ChIP-seq)	<b>Protein structure [14-23]</b>
	features from genomic sequence	1-dimensional structural properties contact map structure model quality assessment
	position specific scoring matrix (PSSM)	<b>Gene expression regulation [24-31]</b>
	physicochemical properties (steric parameter, volume)	splice junction
	Atchley factors (FAC)	genetic variants affecting splicing
	1-dimensional structural properties	sequence specificity
	contact map (distance of amino acid pairs in 3D structure)	<b>Protein classification [32-33]</b>
	microarray gene expression	super family subcellular localization
		<b>Anomaly classification [34]</b>
		Cancer
<b>Biomedical imaging</b>	magnetic resonance imgae (MRI)	<b>Anomaly classification [41-51]</b>
	radiographic image	gene expression pattern cancer
	positron emission tomography (PET)	Alzheimer's disease
	histopathology image	schizophrenia
	volumetric electron microscopy image	<b>Segmentation [52-60]</b>
	retinal image	cell structure neuronal structure vessel map brain tumor
	in situ hybridization (ISH) image	<b>Recognition [61-65]</b>
		cell nuclei finger joint anatomical structure
		<b>Brain decoding [66-67]</b>
		behavior
<b>Biomedical signal processing</b>	ECoG, ECG, EMG, EOG	<b>Brain decoding [74-86]</b>
	EEG (raw, wavelet, frequency, differential entropy)	behavior emotion
	extracted features from EEG	<b>Anomaly classification [87-94]</b>
	normalized decay	Alzheimer's disease
	peak variation	seizure sleep stage

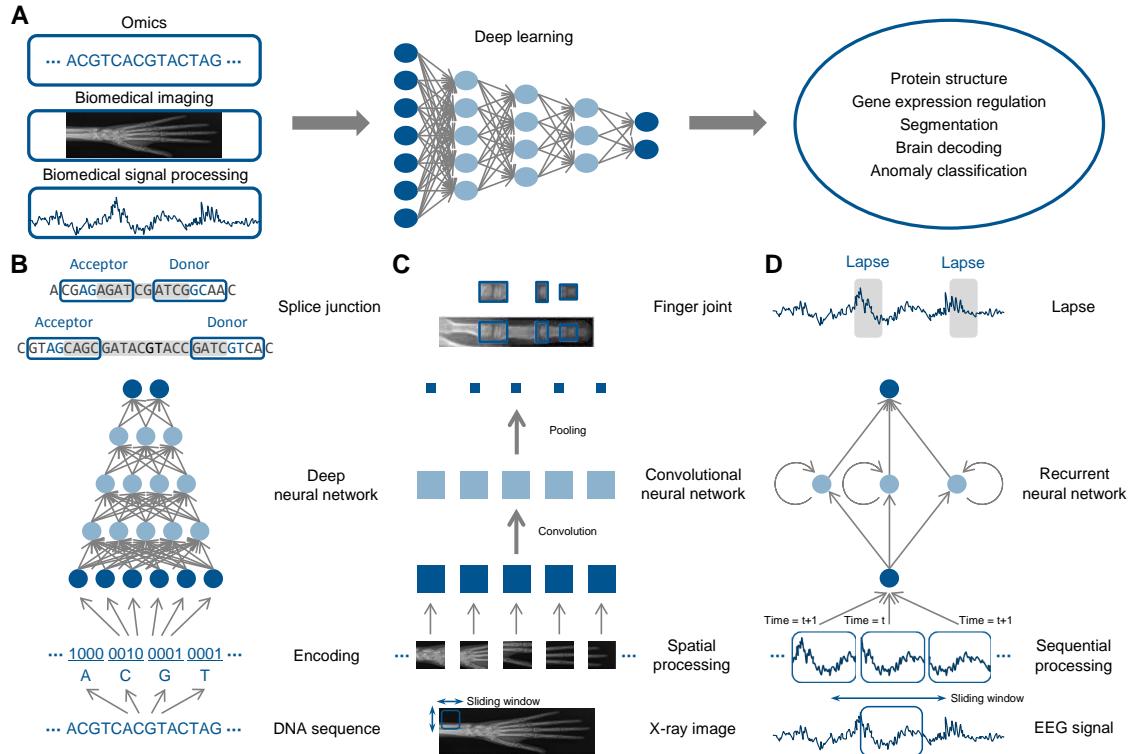
*Omics* is a domain where researchers use genetic information such as genome, transcriptome and proteome to approach problems in bioinformatics. One of the most common input data in omics is the raw form of biological sequences — deoxyribonucleic acid (DNA), ribonucleic acid (RNA), amino acid — which became relatively affordable and easy to obtain with next-generation sequencing technology. Also, extracted features from sequences such as a position

specific scoring matrix (PSSM) [8], physicochemical properties [9, 10], Atchley factors (FAC) [11], and one-dimensional structural properties [12, 13] are often used as input of deep learning algorithms to alleviate difficulties from complex biological data and improve the results. Besides, a protein contact map, which presents the distances of amino acid pairs in their three-dimensional structure, and microarray gene expression data are also used according to the problem characteristics. We categorized the problems in omics into four groups as in Table 1. One of the most researched problems is protein structure, which aims to predict the secondary structure or contact map of the proteins [14-23]. Gene expression regulation [24-31] regarding splice junctions or RNA binding proteins and protein classification [32, 33] regarding super family or subcellular localization are also actively conducted studies. Furthermore, anomaly classification [34] has been approached with omics data to detect cancer.

*Biomedical imaging* is also an actively researched domain since deep learning has been widely used in general image-related tasks. Most of the biomedical images that doctors use to treat patients in real life — magnetic resonance image (MRI) [35, 36], radiographic image [37, 38], positron emission tomography (PET) [39], histopathology image [40] — have been used as input data of deep learning algorithms. We categorized the problems in biomedical image into four groups as in Table 1. One of the most researched problems is anomaly classification [41-51] to diagnose diseases such as cancer or schizophrenia. Just like general image-related tasks, segmentation [52-60] regarding partitioning specific structures such as cell structure or brain tumor and recognition [61-65] regarding detection of cell nuclei or finger joint are studied a lot in biomedical image. Specifically, head MRIs have also been used in brain decoding [66, 67] to interpret people's behavior or emotion.

*Biomedical signal processing* is a domain where researchers use recorded electrical activity of the human body to approach the problems in bioinformatics. Various data exist such as electroencephalography (EEG) [68], electrocorticography (ECoG) [69], electrocardiography (ECG) [70], electromyography (EMG) [71], and electrooculography (EOG) [72, 73], but so far, most studies are concentrated in EEG. Because recorded signals are usually noisy and include many artefacts, raw signals are often decomposed into wavelet or frequency components before they are used as an input in deep learning algorithms. Also, human-designed features like normalized decay and peak variation are used in some studies to improve the results. We categorized the problems in biomedical signal into two groups as in Table 1. Brain decoding

[74-86] using EEG signals and anomaly classification [87-94] to diagnose diseases are the most researched problems in the domain.



**Figure 1:** Application of deep learning in bioinformatics research. (A) Overview diagram with input data and research objectives. (B) A research example in the omics domain. Prediction of splice junctions in DNA sequence data with deep neural network [25]. (C) A research example in biomedical imaging. Finger joint detection from X-ray images with convolutional neural network [63]. (D) A research example in biomedical signal processing. Lapse detection from EEG signal with recurrent neural network [94].

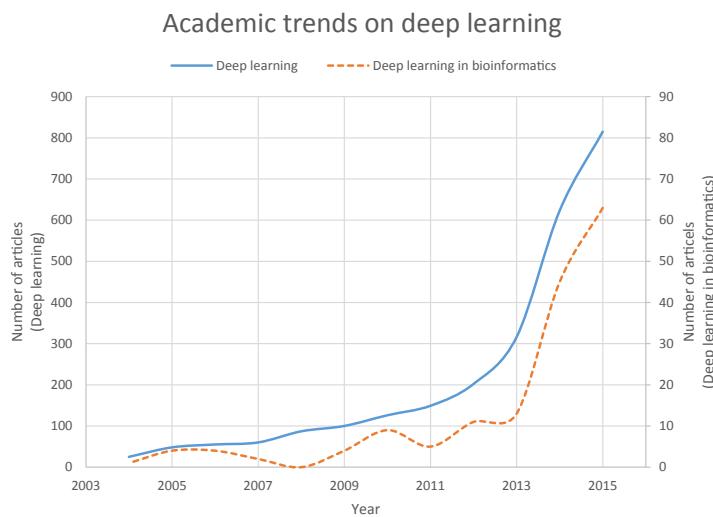
This article reviews the deep learning research in bioinformatics. The goal of this article is to provide valuable insights for researchers to apply deep learning in their bioinformatics research. To the best of our knowledge, we are the first to review the deep learning algorithms in bioinformatics.

This article is organized as follows. ‘Deep learning: a quick look’ section presents an introduction of deep learning, including its advantages, brief descriptions of well-known architectures, and a few available software online. The following sections of ‘Deep neural network’, ‘Convolutional neural network’, ‘Recurrent neural network’, and ‘Modified neural network’ focus on the explanations of each architecture and their applications in bioinformatics

research. ‘Discussion’ introduces some problems that may occur and future research directions. Finally ‘Conclusion’ section summarizes the paper.

## Deep learning: a quick look

Deep learning is a type of machine learning algorithm that uses an artificial neural network of multiple nonlinear layers. The key aspect of deep learning is that suitable features are not designed by human engineers but learned from the data themselves. As one of the representation learning methods, deep learning can learn and discover hierarchical representations of data with increasing level of abstraction [7]. For example, in image recognition, it can be interpreted that feature learning is done in the order of pixel, edge, texton, motif, part, and object. Similarly, in text recognition, features are learned in the order of character, word, word group, clause, sentence, and story [95].



**Figure 2:** Published deep learning articles by year. The number of articles is based on the search results on <http://www.scopus.com> with the two queries: “Deep learning,” “Deep learning,” and “bio.”\*

Research in deep learning, and specifically in bioinformatics, has been rapidly increasing since early 2000s, as in Figure 2, and made breakthroughs in various fields where the artificial intelligence community was struggling for many years [7]. The greatest advancement so far has been made in image and speech recognition [96-102], and deep learning is also showing promising results in natural language processing [103, 104] and language translation [105].

Several deep learning architectures exist and are used according to input data characteristics and research objectives. As in Table 2, we categorized deep learning architectures into four groups: deep neural network (DNN) [106-110], convolutional neural network (CNN) [111-113], recurrent neural network (RNN) [114-118], and modified neural network (MNN) [21, 119-121]. In some paper, DNN often refers to the entire deep learning architectures. However in this review, we use the term “DNN” to specifically refer to multilayer perceptron (MLP) [106], stacked auto-encoder (SAE) [107, 108], and deep belief network (DBN) [109, 110] which uses perceptron [122], auto-encoder (AE) [123], and restricted Boltzmann machine (RBM) [124, 125] as a building block of neural networks, respectively. CNN is an architecture that especially succeeded in image recognition and consists of convolution layers, nonlinear layers, and pooling layers. RNN is designed to utilize sequential information of input data by having cyclic connections among building blocks like perceptron, long short-term memory (LSTM) [116, 117], or gated recurrent unit (GRU) [118]. In addition, many other modified deep learning architectures have been suggested such as deep spatio-temporal neural network (DST-NN) [21], multidimensional recurrent neural network (MD-RNN) [119], and convolutional auto-encoder (CAE) [120, 121]. So in this review, we refer to them as MNN.

**Table 2:** Categorization of deep learning applied research in bioinformatics

	Omics		Biomedical imaging		Biomedical signal processing	
	Research topics	Reference	Research topics	Reference	Research topics	Reference
<b>Deep neural network</b>	Protein structure	[14-17]	Anomaly classification	[41-43]	Brain decoding	[74-79]
	Gene expression regulation	[24-27]	Segmentation	[52]	Anomaly classification	[87-91]
	Anomaly classification	[32]	Brain decoding	[66-67]		
			Recognition	[61]		
<b>Convolutional neural network</b>	Gene expression regulation	[28-30]	Anomaly classification	[44-51]	Brain decoding	[80-83]
			Segmentation	[53-59]	Anomaly classification	[92]
			Recognition	[62-65]		
<b>Recurrent neural network</b>	Protein structure	[18-20]			Brain decoding	[84]
	Protein classification	[32-33]			Anomaly classification	[93-94]
	Gene expression regulation	[31]				
<b>Modified neural network</b>	Protein structure	[21-23]	Segmentation	[60]	Brain decoding	[85-86]

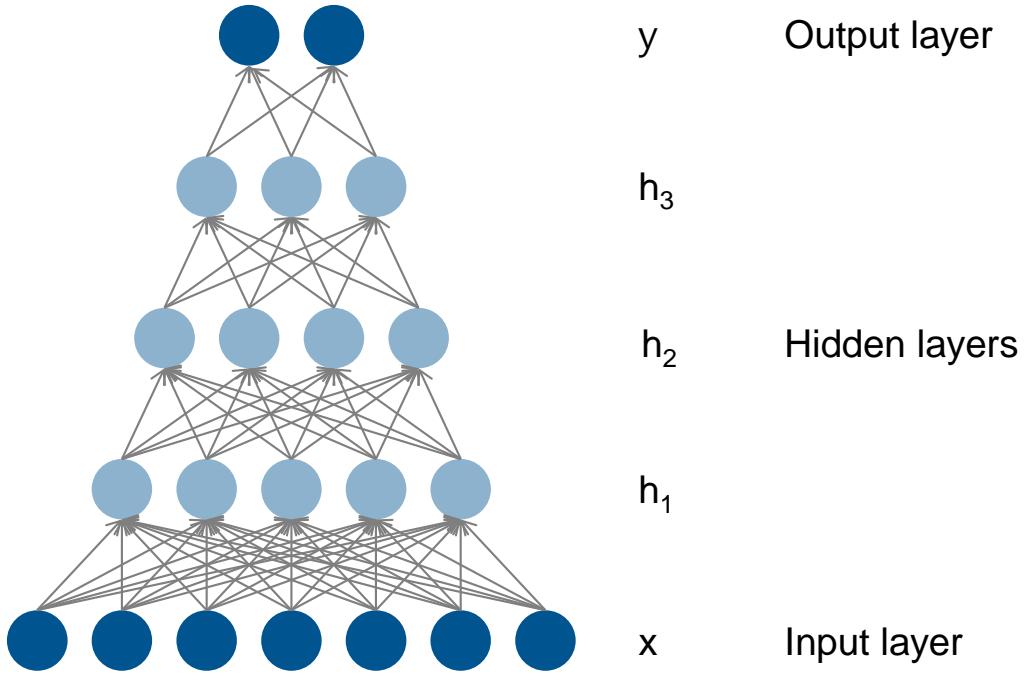
In order to actually implement deep learning algorithms, it requires a great deal of effort in algorithmic and optimization details. Fortunately, there exist many available deep learning software online. The most well-known software are the C++-based Caffe [126], MATLAB-based DeepLearnToolBox [127], LuaJIT-based Torch7 [128], and python-based Theano [129, 130]. Also, Keras [131] and Pylearn2 [132], which provide more convenient interface based on Theano, are widely used. A variety of software including recently released TensorFlow [133] are constantly developed and complemented.

## Deep neural network

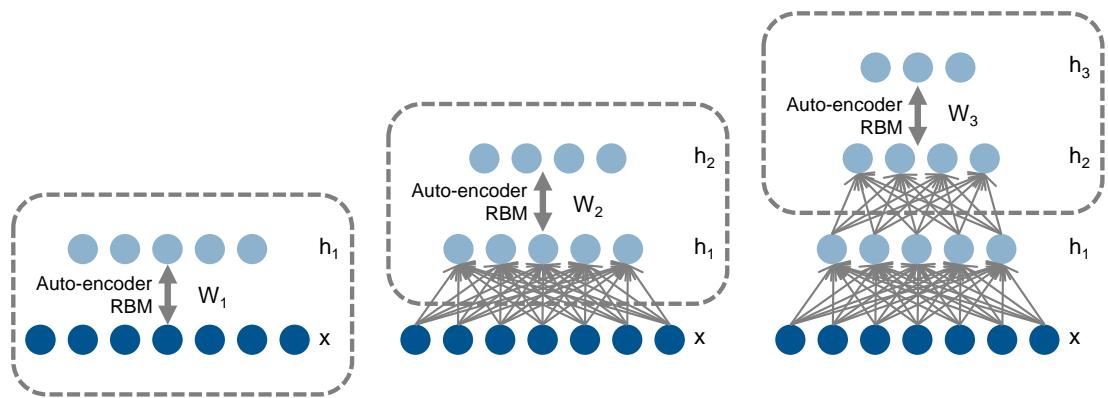
### Introduction

The basic structure of DNN consists of an input layer, multiple hidden layers, and an output layer as in Figure 3. Once input data are given to the DNN, output values are sequentially computed along the layers of the network. First, at each layer, the input vector, which consists of output values of each unit in the layer below, is multiplied by the weight vector for each unit in the current layer producing the weighted sum. Then a nonlinear function such as sigmoid, hyperbolic tangent, or rectified linear unit (ReLU) [134] is applied to the weighted sum to compute the output values of the layer. Through the computation in each layer, the representations in the layer below are transformed into slightly more abstract representations [7]. Therefore, training of DNN aims to optimize the weight vectors so that the most suitable representations could be learned. Based on the types of layer used in DNN and the corresponding learning method, DNN can be classified as MLP, SAE, and DBN.

MLP has a similar structure as the usual artificial neural network, except that more layers are stacked. It is trained in a purely supervised manner that uses only labeled data by initializing the parameters randomly and then training with backpropagation [135] and stochastic gradient descent (SGD) [136]. Since the training method is a process of optimization in high dimensional parameter space, MLP is usually used when a huge number of labeled data are available [95].



**Figure 3:** Basic structure of DNN with input units  $x$ , three hidden units  $h_1$ ,  $h_2$ , and  $h_3$ , in each layer and output units  $y$  [106]. At each layer, weighted sum and nonlinear function of its inputs are computed so that the hierarchical representations can be obtained.



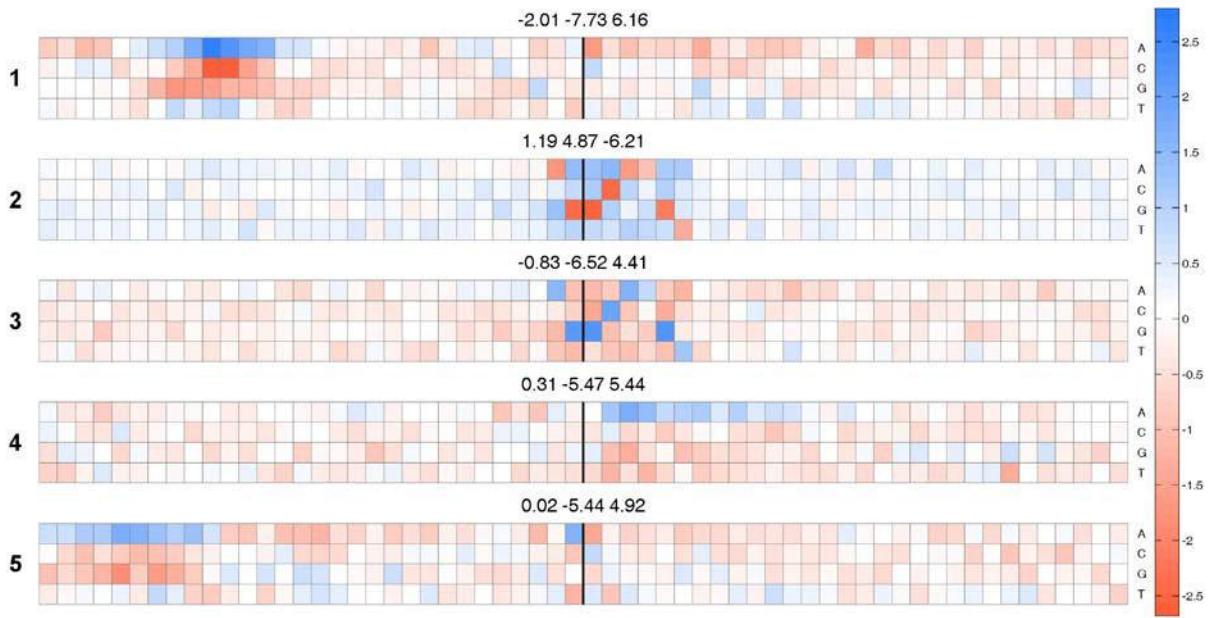
**Figure 4:** Unsupervised layer-wise pre-training process in SAE and DBN [109]. First, train weight matrix  $W_1$  between input units  $x$  and hidden units  $h_1$  in the first hidden layer as an RBM or AE. After the  $W_1$  is trained, another hidden layer is stacked, and the obtained representations in  $h_1$  are used to train  $W_2$  between hidden units  $h_1$  and  $h_2$  as another RBM or AE. The process is repeated for the desired number of layers.

SAE and DBN use AE and RBM as a building block of the architectures, respectively. The main difference between the previous MLP is that training is done in two phases of unsupervised pre-training and supervised fine-tuning. First, in unsupervised pre-training as in Figure 4, each layer is stacked one at a time and trained layer-wise as an AE or RBM using unlabeled data. Afterward, in supervised fine-tuning, an output classifier layer is stacked, and the whole neural network is optimized by retraining with labeled data. Since both SAE and DBN exploit the unlabeled data and can be a great help to avoid overfitting, researchers are able to obtain regularized results even when labeled data are insufficient as in many problems in the real world [137].

## **Deep neural network in bioinformatics**

### *Omics*

DNN has been widely applied in protein structure prediction [14-17] problems. Since the complete prediction in three-dimensional space is such a complex and challenging task, several studies used simpler approaches such as a predicting secondary structure or a torsion angle of protein. For instance, in Heffernan et al. [15], SAE was applied to protein amino acid sequences in the prediction problems of secondary structure, torsion angle, and accessible surface area. Another example is Spencer et al. [16] where DBN was applied to amino acid sequences along with the PSSM and FAC features in the protein secondary structure prediction. Gene expression regulation [24-27] is another research topic that DNN showed great capabilities. For example, Lee et al. [25] utilized DBN in splice junction prediction, which is one of the major research problems in understanding gene expression [138]. The paper proposed a new DBN training method called boosted contrastive divergence for imbalance data and a new regularization term for sparsity of DNA sequences and then showed not only significantly improved performance but also the ability to detect subtle non-canonical splicing signals as in Figure 5. Also, Chen and Li et al. [27] applied MLP on both microarray and RNA-seq based gene expression data to infer expression of target genes up to 21000, only using the approximately 1000 land mark genes. In terms of anomaly classification, in Fakoor et al. [34], principal component analysis (PCA) [139] for dimensionality reduction was used on microarray gene expression data, and then SAE was applied to classify various cancers like acute myeloid leukemia, breast cancer, and ovarian cancer.



**Figure 5:** Non-canonical splicing sites identified in Lee et al. [25]. The three numbers above each matrix represent the discriminative scores for acceptor, non-boundary and donor. The matrixes are colored according to weight values so that a darker blue represents a higher positive value. It reveals novel non-canonical splicing patterns such as GCA or NAA at splicing boundaries and contiguous A's in exon regions near the splicing boundaries.

### *Biomedical imaging*

In terms of biomedical imaging, DNN has been applied in several research areas including anomaly classification [41-43], segmentation [52], recognition [61], and brain decoding [66, 67]. Regardless of the various image sources, Plis et al. [41] classified schizophrenia patients using DBN on brain MRIs, and Xu et al. [61] used SAE in cell nuclei detection from histopathology images. Also interestingly, similar to the research of handwritten digit image recognition, Van Gerven et al. [66] classified handwritten digit images with DBN not by analyzing the images themselves but by indirectly analyzing functional MRIs of participants who are looking at the digit images.

### *Biomedical signal processing*

Since biomedical signal usually contains a lot of noise and artefacts, decomposed features are more frequently used rather than the raw signal. For brain decoding [74-79], An et al. [75] applied DBN on frequency components of EEG to classify left and right hand motor imagery.

Additionally, in Jia et al. [77] and Jirayucharoensak et al. [79], emotion classification was conducted using DBN and SAE, respectively. For anomaly classification [87-91] problems, Huanhuan et al. [87], which is one of the few studies to apply DBN on ECG signals, classified each bit into either normal or abnormal beat. There are also few studies that used raw EEG signals. Wulsin et al. [88] analyzed individual second-long waveform abnormality using DBN on both raw EEG signal and extracted features as input, whereas Zhao et al. [90] only used raw EEG signal as input to DBN in diagnosing Alzheimer's disease.

## **Summary**

DNN is renowned for its suitability in the analysis of internal correlations in high dimensional data. Given that bioinformatics data are typical complex and high dimensional data, we could look into diverse studies applying DNN over omics, biomedical imaging, and biomedical signal processing. However, it occurred to us that the capabilities of DNN are not yet fully exploited. Although the key aspect of DNN is that hierarchical features are learned from data, human designed features were often given as input instead of raw data forms. We expect that future progress of DNN in bioinformatics will come from researching proper ways to encode the raw data forms and learn suitable features from the raw forms.

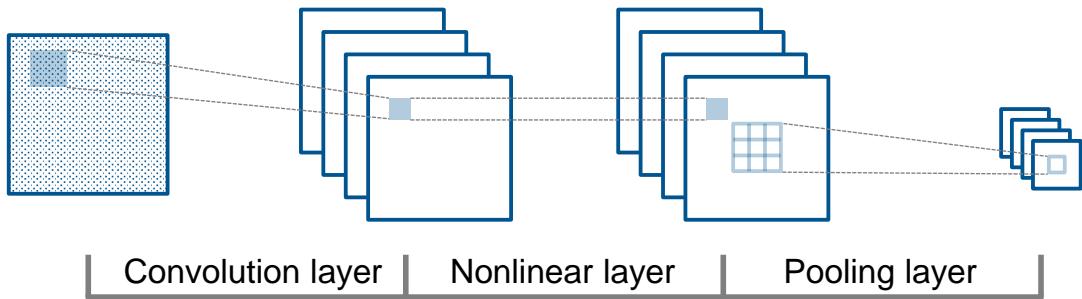
# **Convolutional neural network**

## **Introduction**

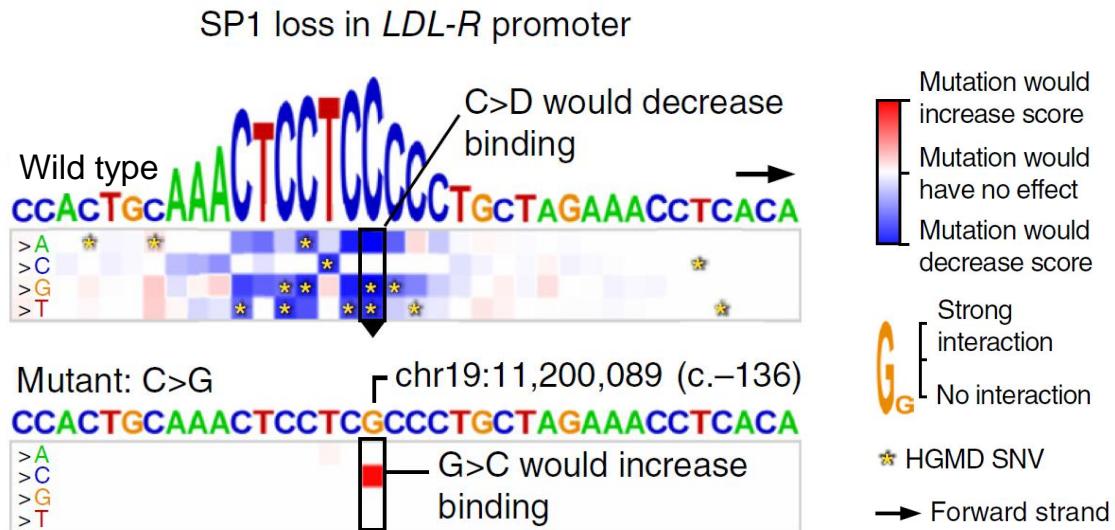
CNN, designed to process multiple array types of data, especially two-dimensional images, is directly inspired by the visual cortex of the brain. In the visual cortex, there exists a hierarchy of two basic cell types: simple cells and complex cells [140]. Simple cells react to primitive patterns in sub-regions of visual stimuli beforehand. Then complex cells put together the information from simple cells and identify more intricate forms. Since the visual cortex is such a powerful and natural visual processing system, CNN applied three key ideas to imitate: local connectivity, invariance to location, and invariance to local transition [7].

The basic structure of CNN consists of convolution layers, nonlinear layers, and pooling layers as in Figure 6. To utilize highly correlated sub-regions of data, at each convolution layer, groups of local weighted sum called feature maps are obtained by computing convolutions between local patches and weight vectors called filters. Furthermore, since identical patterns can appear

regardless of the location in the data, filters are applied repeatedly across the entire data, which also gives advantage in training efficiency by reducing the number of parameters to learn. Then nonlinear layers increase the nonlinear properties of feature maps. At each pooling layer, max or average subsampling of non-overlapping regions in feature maps is carried out. The non-overlapping subsampling enables CNN to handle somewhat different but semantically similar features and aggregate local features to identify more complex features.



**Figure 6:** Basic structure of CNN consisting of convolution layer, nonlinear layer, and pooling layer [112]. The convolution layer of CNN uses multiple learned filters to obtain multiple filter maps detecting low-level filters, and then the pooling layer combines them into higher-level features.



**Figure 7:** Analysis of disease-associated genetic variants affecting transcription factor binding in Alipanahi et al. [29]. It shows that in a human gene mutation database (HGMD), single nucleotide variants affect an SP1 transcription binding site in LDL-R promoter, causing familial hypercholesterolemia [141, 142].

## **Convolutional neural network in bioinformatics**

### *Omics*

A few studies have been conducted applying CNN to gene expression regulation [28-30] problems. For example, Denas et al. [28] used CNN on ChIP-seq data to analyze gene expression levels. Also, recently in Alipanahi et al. [29], CNN was applied on both microarray and sequencing data of RNA binding proteins to learn sequence binding specificities, and then DNN was used to analyze disease-associated genetic variants affecting transcription factor binding as in Figure 7.

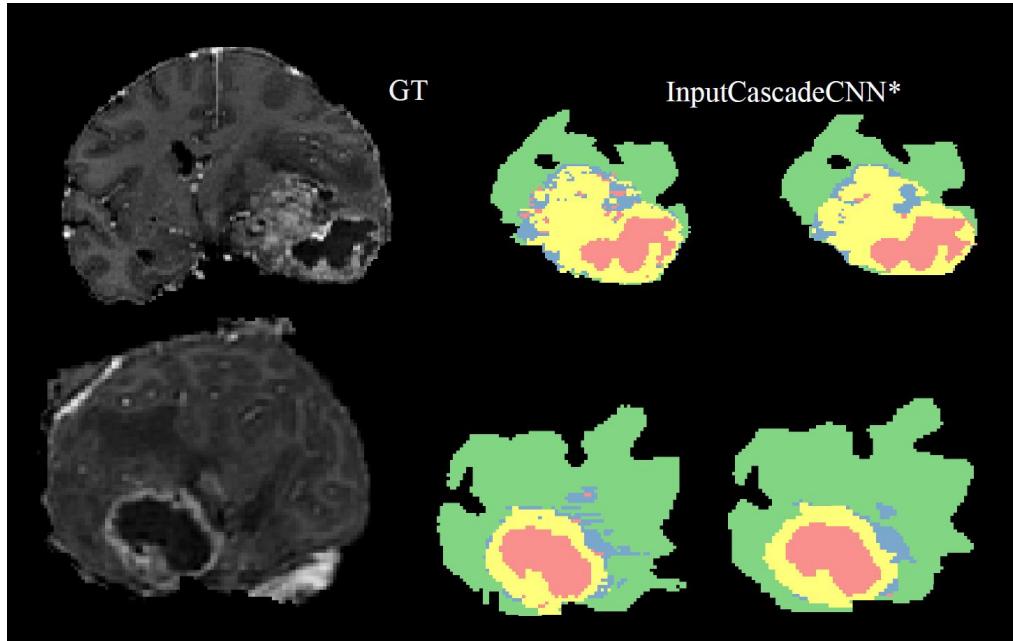
### *Biomedical imaging*

The largest number of research has been conducted in biomedical imaging since the problems have the similar form as general image-related tasks. In anomaly classification [44-51], Roth et al. [44] applied CNN on three different CT image datasets to classify sclerotic metastases, lymph node, and colonic polyp. Additionally, Ciresan et al. [47] used CNN in mitosis detection among breast cancer histopathology images which is crucial in cancer screening and assessment. PET images of esophageal cancer were also utilized in Ypsilantis et al. [48] to predict responses to neoadjuvant chemotherapy. Other applications of CNN can be found in segmentation [53-59] and recognition [62-65] as well. In Ning et al. [53], pixel-wise segmentation of cell wall, cytoplasm, nucleus membrane, nucleus, and outside medium in cell microscopic images was researched. Also, Havaei et al. [58] proposed a cascaded CNN architecture exploiting both local and global contextual features and performed brain tumor segmentation from MRIs as in Figure 8. In terms of recognition, Cho et al. [62] researched anatomical structure recognition among CT images, and Lee et al. [63] proposed a CNN-based finger joint detection system, FingerNet, which is a crucial step for medical examinations of bone age, growth disorders, and rheumatoid arthritis [143].

### *Biomedical signal processing*

Raw EEG signals have been analyzed in brain decoding [80-83] and anomaly classification [92] using CNN, which performs one-dimensional convolutions. For instance, Stober et al. [81] classified the rhythm type and genre of the music that the participants are listening to, and Cecotti et al. [83] classified the characters that the participants are looking at. There is another approach in applying CNN to biomedical signal. In Mirowski et al. [92], features such as phase-

locking synchrony and wavelet coherence were extracted and coded as a color of each pixel which formulates two-dimensional patterns. Then an ordinary two-dimensional CNN, like the one used in biomedical imaging, was used in seizure prediction.



**Figure 8:** Brain tumor segmentation results in Havaei et al. [58]. The images in each column represent original T1 contrast MRIs, ground truth segmentation, and output of the CNN from left to right. Segmentation images are colored so that green, yellow, red, and blue regions indicate edema, enhanced tumor, necrosis, and non-enhanced tumor, respectively.

## Summary

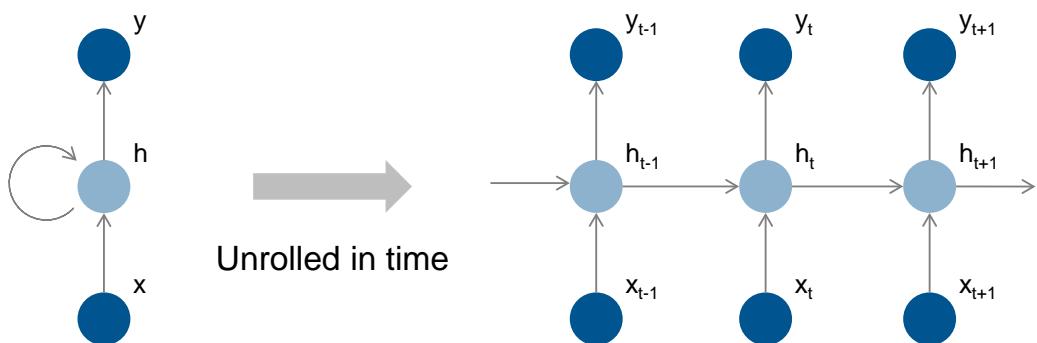
Upon the capabilities of CNN in analyzing spatial information, we could observe that most researches of CNN in bioinformatics are focused on biomedical imaging so far. Intuitively, it is natural that CNN is not the first choice of deep learning architecture in omics and biomedical signal processing since usual data in the domain does not seem to be spatial information. However, two-dimensional data such as interactions between biological sequences and the time-frequency matrix of a biomedical signal can still be considered as spatial information. Thus, we believe that CNN has great potential in the domains and poised to make great impact in the future.

# Recurrent neural network

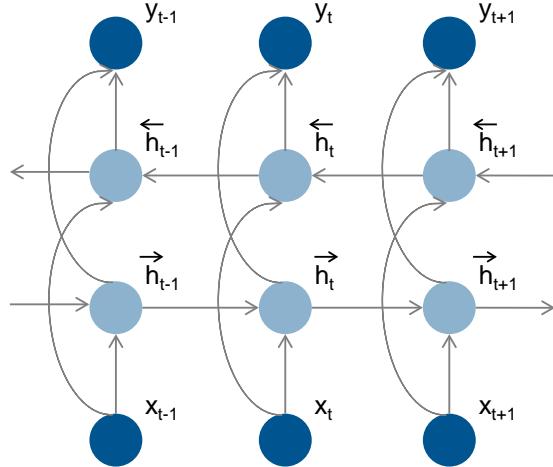
## Introduction

RNN, which is designed to utilize sequential information, has the basic structure of having a cyclic connection as in Figure 9. Since input data are processed sequentially, recurrent computation is carried out in the hidden unit where cyclic connection exists. Therefore, past information is implicitly stored in the hidden unit called state vector, and using the state vector, output for the current input is computed considering the whole past inputs [7]. Since there are many cases that both past and future inputs affect output for the current input, such as in speech recognition, bidirectional recurrent neural network (BRNN) [144] as in Figure 10 has also been designed and used widely.

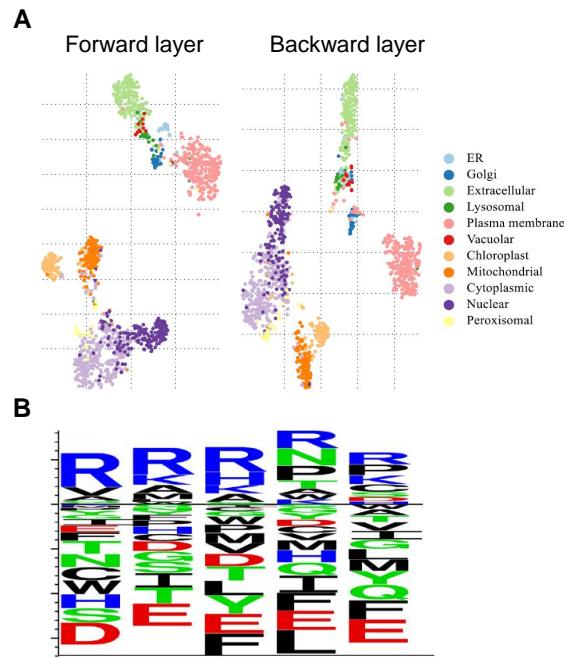
Although RNN does not seem to be deep as DNN or CNN in terms of the number of layers, it can be seen as an even deeper structure if unrolled in time as in Figure 9. Therefore, for a long time, researchers struggled against vanishing gradient problems while training the RNN and had difficulties in learning long-term dependency among data [115]. Fortunately, researchers showed that substituting the simple perceptron hidden units to more complex ones, LSTM [116, 117] or GRU [118], which function as memory cell significantly helps to prevent problems, and lately, RNN is successfully used in many areas including natural language processing [103, 104] and language translation [105, 118].



**Figure 9:** Basic structure of RNN with an input unit  $x$ , a hidden unit  $h$ , and an output unit  $y$  [7]. A cyclic connection exists so that the computation in the hidden unit gets inputs from the hidden unit at previous time step as well as the input unit at the current time step. The recurrent computation can be expressed more explicitly if the RNN is unrolled in time. The index of each symbol represents the time step. In this way,  $h_t$  receives input from  $x_t$  and  $h_{t-1}$ , and then propagates the computed results to  $y_t$  and  $h_{t+1}$ .



**Figure 10:** Basic structure of BRNN unrolled in time [144]. It contains two hidden units  $\vec{h}_t$  and  $\tilde{\vec{h}}_t$  for each time step.  $\vec{h}_t$  receives input from  $x_t$  and  $\vec{h}_{t-1}$  to reflect past information. On the other hand,  $\tilde{\vec{h}}_t$  receives input from  $x_t$  and  $\tilde{\vec{h}}_{t+1}$  to reflect future information. Then information from both hidden units is propagated to  $y_t$ .



**Figure 11:** Learned amino acid sequence representations in Sønderby et al. [33]. (A) t-SNE plot of hidden representations in forward and backward layers of BNNS. It reveals that proteins from the same subcellular location generally group together. (B) Example of learned convolutional filter representing a nuclear localization signal. It is visualized as a PSSM logo, where the height of each column and letter can be interpreted as position and amino acid importance, respectively.

## **Recurrent neural network in bioinformatics**

### *Omics*

RNN has been expected to be an appropriate deep learning architecture because biological sequences have variable lengths, and their sequential information has a great importance. A few studies have been conducted to apply RNN in protein structure prediction [18-20], protein classification [32, 33], and gene expression regulation [31] problems. In the early studies, Baldi et al. [18] used BRNN with perceptron hidden units in protein secondary structure prediction. Then after it became widely known that LSTM hidden units show better performance, Sønderby et al. [33] applied BRNN with the LSTM hidden units and a 1D convolution layer to learn representations from amino acid sequences and classify subcellular locations of proteins as in Figure 11. Furthermore, Lee et al. [31] exploited RNN with LSTM hidden units in splice junction prediction, and obtained significant increase of accuracy with respect to the state-of-the-art DBN approach [25] showing the great capabilities of RNN in analyzing DNA sequences.

### *Biomedical imaging*

Traditionally, images are considered as the data which involve internal correlations or spatial information rather than sequential information. Treating the biomedical images as non-sequential data, most studies in the domain have chosen the approaches regarding DNN or CNN instead of RNN. However there are still some studies trying to apply unique capabilities of RNN in image data using modified RNN structure, MD-RNN [119]. These studies will be discussed in the next section with a more detailed explanation of the modified structure.

### *Biomedical signal processing*

Since biomedical signal is naturally sequential data, RNN is a quite proper deep learning architecture to analyze the data and has been expected to produce promising results. To present some of the studies in brain decoding [84] and anomaly classification [93, 94], Petrosian et al. [93] applied perceptron RNN on raw EEG signal and its wavelet decomposed features in seizure prediction. In Davidson et al. [94], LSTM RNN was used on EEG log-power spectrum features in lapse detection.

## **Summary**

Although RNN in bioinformatics is still in the early stages compared with DNN and CNN, its capabilities in analyzing sequential information create such a high expectation. Not only research in terms of typical sequential data in omics and biomedical signal processing has not been fully explored, but there also exist a lot of areas in which RNN has great potentials. Analysis of dynamic CT and MRI [145, 146] consisting multiple sequential images are one of the areas, and even though we have not focused in this review, biomedical text analysis [147] such as electronic medical records and research papers will be able to make major progress with RNN.

## Modified neural network

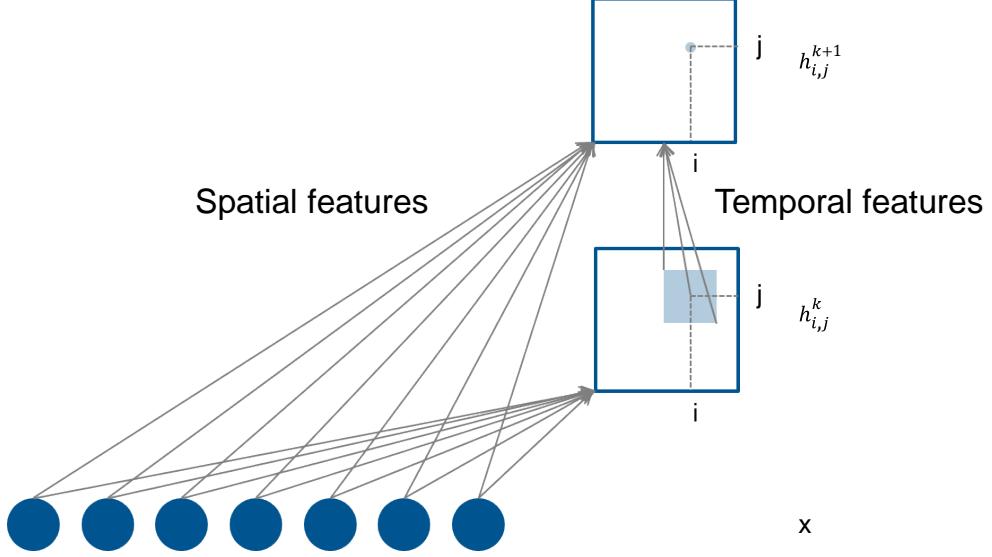
### Introduction

MNN refers to modified deep learning architectures besides DNN, CNN, and RNN. In this paper, we introduce three MNNs — DST-NN, MD-RNN, and CAE — and their applications in bioinformatics.

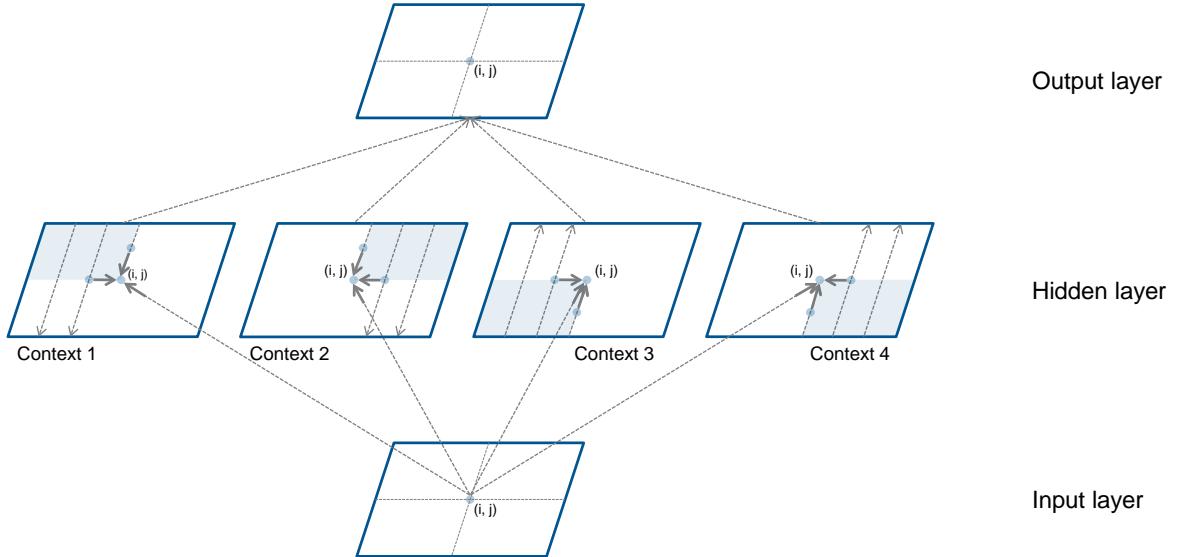
DST-NN [21] is designed to learn multi-dimensional output targets through progressive refinement. The basic structure of DST-NN consists of multi-dimensional hidden layers as in Figure 12. The key aspect of the structure, progressive refinement, considering the local correlations is done via input feature compositions in each layer: spatial features and temporal features. Spatial features refer to the original input to the whole DST-NN and are used identically in every layer. On the other hand, temporal features refer to gradually altered features so as to progress to the upper layers. Except for the first layer, to compute each hidden unit in the current layer, only the adjacent hidden units of the same coordinate in the layer below are used so that the local correlations are reflected progressively.

MD-RNN [119] is designed to apply the capabilities of RNN to non-sequential multi-dimensional data by treating them as groups of sequential data. For instance, two-dimensional data are treated as groups of horizontal and vertical sequence data. The basic structure of MD-RNN is the same as Figure 13. Similar to BRNN which uses contexts in both directions in the one-dimensional data, MD-RNN uses contexts in all possible directions in the multi-dimensional data. In the example of two-dimensional data, four contexts which vary depending on the order of processing the data are reflected in the computation of four hidden units for

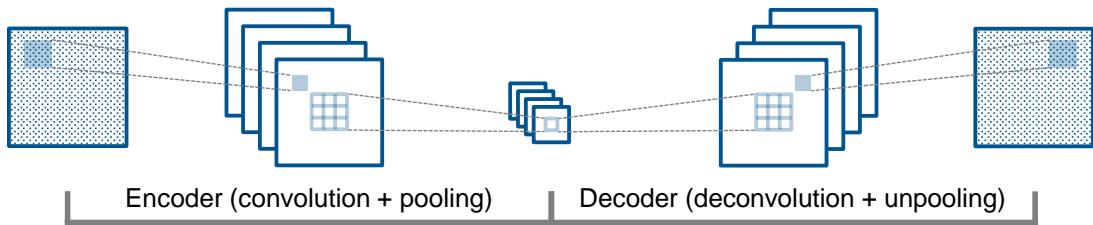
each position in the hidden layer. Then the hidden units are connected to a single output layer, and final results are computed upon consideration of all the contexts.



**Figure 12:** Basic structure of DST-NN [21]. The notation  $h_{i,j}^k$  represents hidden unit in  $(i, j)$  coordinate of the  $k$ -th hidden layer. To conduct the progressive refinement, the neighborhood units of  $h_{i,j}^k$  as well as input units  $x$  are used in the computation of  $h_{i,j}^{k+1}$ .



**Figure 13:** Basic structure of MD-RNN for two-dimensional data [119]. It contains four groups of two-dimensional hidden units, each reflecting different contexts. For example,  $(i, j)$  hidden unit in the context 1 group receives input from  $(i-1, j)$  and  $(i, j-1)$  hidden units in the context 1 as well as  $(i, j)$  unit from input layer so that the upper-left information is reflected. Then the hidden units from all four contexts are propagated to compute  $(i, j)$  unit in the output layer.



**Figure 14:** Basic structure of CAE consisting of convolution and pooling layer working as an encoder, and deconvolution and unpooling layer working as a decoder [121]. The basic idea is similar as the AE that learns hierarchical representations through reconstructing its input data, but CAE additionally utilizes spatial information by integrating convolutions.

CAE [120, 121] is designed to utilize the advantages of both AE and CNN so that it can learn good hierarchical representations of data reflecting spatial information and well regularized by unsupervised training. The basic structure of CAE is the same as Figure 14, and the idea behind CAE is as follows: In the training of AE, reconstruction error is minimized using encoder and decoder, which, respectively, extracts feature vectors from input data and recreates the data from the feature vectors. Meanwhile in CNN, convolution and pooling layers can be seen as some sort of encoder. Therefore, the CNN encoder and decoder consisting of the deconvolution and unpooling layer are integrated to form CAE and trained in the same manner as AE.

## Modified neural network in bioinformatics

### *Omics*

MNN has been used in protein structure [21-23] research, specifically contact map prediction. Di Lena et al. [22] applied DST-NN using spatial features of protein secondary structure, orientation probability, alignment probability, and many others. Additionally, in Baldi et al. [23], MD-RNN was applied on amino acid sequences, correlated profiles, and protein secondary structures.

### *Biomedical imaging*

In terms of biomedical imaging, MNN, specifically MD-RNN, has been applied beyond two-dimensional images to three-dimensional images. In Stollenga et al. [60], MD-RNN was applied on three-dimensional electron microscopy images and MRIs to segment neuronal structures.

## *Biomedical signal processing*

CAE has been applied in brain decoding [85, 86] research. In Wang et al. [85], finger flex and extend classification was carried out using raw ECoG signals. Also, Stober et al. [86] classified the rhythm type of the music that the participants are listening to with raw EEG signals.

## **Summary**

As deep learning is a rapidly growing research area, a lot of new deep learning architectures are being suggested yet to be widely applied in bioinformatics. Newly proposed architectures usually have different potentials from the existing ones, so we expect them to produce promising results in the areas where the existing ones have not succeeded yet. Besides, the MNN discussed in the review, we believe that the recently emerging neural Turing machine [148] which can learn algorithms and ladder network [149] and can take advantage of both unsupervised and supervised learning will become more important in the future.

## **Discussion**

### **Limited size and class imbalance data**

The most frequent problems in the application of deep learning in bioinformatics are limited size and class imbalance data. Since the optimization of a tremendous number of weight parameters in the neural network according to data and research objectives is the key aspect of deep learning, it inevitably requires enormous training data [150]. At the same time, if training data are class imbalance, it is even more difficult to properly train the neural network [151, 152]. The standard performance measures used during training such as accuracy rate are often biased toward the majority class. For example, in a class imbalance classification problem of A and B constituting 99% and 1%, respectively, deep learning algorithms are less likely to learn suitable features since 99% accuracy can be achieved by simply classifying all the data into A. Furthermore, a small number of minority class examples are often considered as outliers, and even little noise in the data can significantly degrade the identification.

Unfortunately, limited size and class imbalance data are common problems in bioinformatics [153]. Biomedical data are limited in many cases because data acquisition processes are usually complex and expensive. Furthermore, if they are disease related, not only the number of

patients are small in the first place, but also data are rarely disclosed to the public due to privacy restrictions [154]. Even when there is a relatively large number of data, usually, biomedical data are class imbalance in their nature such as in splice junctions in omics [155].

Approaches to the problems of limited size and class imbalance data can be divided into two categories concerning preprocessing and training [152]. First, in the preprocessing approaches, oversampling and undersampling are widely used to enrich and rebalance the data. Oversampling replicates or creates new data from the whole data or minority class. Meanwhile, undersampling selects some data from the majority class. For example, Li et al. [46] and Roth et al. [64] carried out enrichment of CT images by spatial deformations such as random shifting and rotation. Since there is a limitation in learning from complex data when data size is limited, human-designed features are extracted and often used instead of raw data forms. Research in omics and biomedical signal processing that used the human designed features as input data such as PSSM or wavelet components can be understood in the same context.

Among the training approaches, pre-training is widely used to deal with limited size and class imbalance data. With the unsupervised pre-training with RBM or AE, it can be a great help to prevent overfitting and produce more regularized results especially when the data size is limited [137]. Also, transfer learning, which consists of two steps, pre-training with sufficient data from similar but different domains and fine-tuning with the real data, is also widely studied [156]. For instance, Bar et al. [51] carried out pre-training of CNN with ImageNet database of natural images [157] and fine-tuning with chest X-ray images to identify chest pathologies and classify healthy and abnormal images. Besides pre-training, some sophisticated training methods are researched as well. Lee et al. [25] suggested DBN with boosted categorical RBM, and Havaei et al. [58] suggested CNN with two-phase training which combined ideas of undersampling and pre-training.

## **Changing the black-box into the white-box**

One of the main criticisms against deep learning is that it is used as a black-box. In other words, even though it produces outstanding results, we know very little about how it gives such results internally. In bioinformatics, especially in biomedical domains, it is absolutely not enough to just give good outcomes. Since many studies are connected to patients' health, it is crucial to provide logical reasoning as doctors do in medical treatments currently.

Research of transforming deep learning from the black-box into the white-box is still in the early stages. Nevertheless, one of the most widely used approaches is interpretation through visualizing trained deep learning models. With regard to CNN, deconvolution network has been suggested and successfully visualized trained hierarchical representations for image classification [158]. Additionally, when it comes to RNN, activation values of each LSTM unit from trained character-level language model was visualized and demonstrated interpretable LSTM units which identified high-level patterns such as line lengths and brackets [159]. Besides interpretation through visualization, attention mechanisms [160, 161] which explicitly learn to focus on salient objects and the mathematical rationale behind deep learning [162, 163] are being studied as well.

## **Choosing the appropriate deep learning architecture and hyperparameters**

Choosing the appropriate deep learning architecture is also one of the important problems in the application of deep learning. In order to accomplish good results, it is essential to be well aware of capabilities of each deep learning architecture and select the one according to the capabilities in addition to input data characteristics and research objectives. However, so far the advantages of each architecture are only roughly understood such as DNN is suitable for analysis of internal correlations in high dimensional data, CNN is suitable for analysis of spatial information, and RNN is suitable for analysis of sequential information [164]. Detailed methodology in choosing the most appropriate deep learning architecture still remains as a challenge to be studied in the future.

Even if a deep learning architecture is decided, there are still so many hyperparameters — number of layers, number of hidden units, weight initialization values, learning iteration, learning rate — for researchers to decide that have great influence on the results [165]. In many cases, optimizing hyperparameters has been up to human machine learning experts. However, automation of machine learning (AutoML) research which target the automation of machine learning without expert knowledge is growing constantly [166, 167].

## **Multimodal deep learning**

Multimodal deep learning [168], which exploits information from multiple input sources, is one of the highlighted research areas as the future of deep learning. Bioinformatics, in particular, is expected to benefit greatly since it is a field where various types of data can be utilized naturally [169]. For example, not only omics data, image, signal, drug response,

electronic medical records, and so on are available as input data for the research, but even a single image can also be in many forms such as X-ray, CT, MRI, and PET.

There are already a few studies in bioinformatics using multimodal deep learning. In Suk et al. [43], Alzheimer's disease classification was studied using cerebrospinal fluid (CSF) and brain images in the forms of MRI and PET scan. Also, Soleymani et al. [84] conducted an emotion detection research with both EEG signal and face image data.

## **Accelerating deep learning**

It is well known that the more deep learning model parameters and the more training data are utilized, the better learning performances can be achieved. However, at the same time, it inevitably leads to a drastic increase of training time emphasizing necessity for accelerating deep learning [95, 164].

Approaches to accelerating deep learning can be divided into three groups: advanced optimization algorithms, parallel and distributed computing, and specialized hardware. Since the main reason for the long training time is that optimization of parameters thorough plain SGD takes too long, several studies have been focused on advanced optimization algorithms [170]. Some of the widely employed algorithms include momentum [171], Adagrad [172], batch normalization [173], and Hessian-free optimization [174]. Parallel and distributed computing has showed significant speedups and made possible the practical deep learning research [175-179]. It exploits both scale-up methods using a graphic processing unit (GPU) and scale-out methods in a distributed environment using large-scale clusters of machines. A few deep learning frameworks including the recently released DeepSpark [180] and TensorFlow [133] provide parallel and distributed computing. Although a specialized hardware for deep learning is still in the early stages, it will provide major accelerations and become far more important in the long term [181]. Currently, field programmable gate array (FPGA) based processors are under development, and neuromorphic chips modeled on brain are greatly anticipated as a promising technology [182-184]

## **Conclusion**

Entering the major era of big data, deep learning is taking center stage under international academic and business interests. In bioinformatics where great advances have been made with

conventional machine learning algorithms, deep learning is also highly expected to produce promising results. In this paper, we presented an extensive review of bioinformatics research applying deep learning and looked into it in terms of input data, research objectives, and characteristics of widely used deep learning architectures. For an informative review, we further discussed issues related to problems that might occur and future research directions.

Although deep learning sounds endlessly promising, it is not a silver bullet and cannot provide great results when simply applied in bioinformatics. Still, there are many problems to consider such as limited size and class imbalance data, interpretation of deep learning results, and choosing the appropriate architecture and its hyperparameters. Furthermore, to fully exploit the capabilities of deep learning, multimodality and acceleration of deep learning are the promising areas for further research. Thus, we believe that prudent preparations regarding the issues discussed in the review are the key to success in such research. We hope that this review could provide valuable insights and be a starting point for researchers to apply deep learning in their bioinformatics studies.

## Funding

This work was supported by the National Research Foundation (NRF) of Korea grants funded by the Korean Government (Ministry of Science, ICT and Future Planning) [No. 2011-0009963, No. 2014M3C9A3063541]; the ICT R&D program of MSIP/ITP [14-824-09-014, Basic Software Research in Human-level Lifelong Machine Learning (Machine Learning Center)]; SNU ECE Brain Korea 21+ project in 2015; and Samsung Electronics Co., Ltd.

## Acknowledgements

The authors would like to thank Prof. Russ Altman at Stanford University, Prof. Honglak Lee at University of Michigan, and Prof. Kyunghyun Cho at New York University for helpful discussions on applying artificial intelligence and machine learning to bioinformatics.

## References

1. Manyika J, Chui M, Brown B et al. Big data: The next frontier for innovation, competition, and productivity 2011.
2. Ferrucci D, Brown E, Chu-Carroll J et al. Building Watson: An overview of the DeepQA project. *AI magazine* 2010;31(3):59-79.
3. IBM Watson for Oncology. IBM. <http://www.ibm.com/smarterplanet/us/en/ibmwatson/watson-oncology.html>, 2016.
4. Silver D, Huang A, Maddison CJ et al. Mastering the game of Go with deep neural networks and tree search. *Nature* 2016;529(7587):484-9.
5. DeepMind Health. Google DeepMind. <https://www.deepmind.com/health>, 2016.
6. Larrañaga P, Calvo B, Santana R et al. Machine learning in bioinformatics. *Briefings in bioinformatics* 2006;7(1):86-112.
7. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436-44.
8. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology* 1999;292(2):195-202.
9. Ponomarenko JV, Ponomarenko MP, Frolov AS et al. Conformational and physicochemical DNA features specific for transcription factor binding sites. *Bioinformatics* 1999;15(7):654-68.
10. Cai Y-d, Lin SL. Support vector machines for predicting rRNA-, RNA-, and DNA-binding proteins from amino acid sequence. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 2003;1648(1):127-33.
11. Atchley WR, Zhao J, Fernandes AD et al. Solving the protein sequence metric problem.

*Proceedings of the National Academy of Sciences of the United States of America* 2005;102(18):6395-400.

12. Branden CI. Introduction to protein structure. Garland Science, 1999.
13. Richardson JS. The anatomy and taxonomy of protein structure. *Advances in protein chemistry* 1981;34:167-339.
14. Lyons J, Dehzangi A, Heffernan R et al. Predicting backbone C $\alpha$  angles and dihedrals from protein sequences by stacked sparse auto-encoder deep neural network. *Journal of computational chemistry* 2014;35(28):2040-6.
15. Heffernan R, Paliwal K, Lyons J et al. Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. *Scientific reports* 2015;5.
16. Spencer M, Eickholt J, Cheng J. A Deep Learning Network Approach to ab initio Protein Secondary Structure Prediction. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on* 2015;12(1):103-12.
17. Nguyen SP, Shang Y, Xu D. DL-PRO: A novel deep learning method for protein model quality assessment. In: *Neural Networks (IJCNN), 2014 International Joint Conference on*. 2014. p. 2071-8. IEEE.
18. Baldi P, Brunak S, Frasconi P et al. Exploiting the past and the future in protein secondary structure prediction. *Bioinformatics* 1999;15(11):937-46.
19. Baldi P, Pollastri G, Andersen CA et al. Matching protein beta-sheet partners by feedforward and recurrent neural networks. In: *Proceedings of the 2000 Conference on Intelligent Systems for Molecular Biology (ISMB00), La Jolla, CA*. 2000. p. 25-36.
20. Sønderby SK, Winther O. Protein Secondary Structure Prediction with Long Short Term Memory Networks. *arXiv preprint arXiv:1412.7828* 2014.
21. Lena PD, Nagata K, Baldi PF. Deep spatio-temporal architectures and learning for protein structure prediction. In: *Advances in Neural Information Processing Systems*. 2012. p. 512-20.
22. Lena PD, Nagata K, Baldi P. Deep architectures for protein contact map prediction. *Bioinformatics* 2012;28(19):2449-57.
23. Baldi P, Pollastri G. The principled design of large-scale recursive neural network architectures--dag-rnns and the protein structure prediction problem. *The Journal of Machine Learning Research* 2003;4:575-602.
24. Leung MK, Xiong HY, Lee LJ et al. Deep learning of the tissue-regulated splicing code. *Bioinformatics* 2014;30(12):i121-i9.
25. Lee T, Yoon S. Boosted Categorical Restricted Boltzmann Machine for Computational Prediction of Splice Junctions. In: *International Conference on Machine Learning*. Lille, France, 2015. p. 2483-92.
26. Zhang S, Zhou J, Hu H et al. A deep learning framework for modeling structural features of RNA-binding protein targets. *Nucleic acids research* 2015:gkv1025.
27. Chen Y, Li Y, Narayan R et al. Gene expression inference with deep learning. *Bioinformatics*

- 2016(btW074).
28. Denas O, Taylor J. Deep modeling of gene expression regulation in an Erythropoiesis model. In: *International Conference on Machine Learning workshop on Representation Learning*. Atlanta, Georgia, USA, 2013.
  29. Alipanahi B, Delong A, Weirauch MT et al. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature biotechnology* 2015.
  30. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. *Nature methods* 2015;12(10):931-4.
  31. Lee B, Lee T, Na B et al. DNA-Level Splice Junction Prediction using Deep Recurrent Neural Networks. *arXiv preprint arXiv:1512.05135* 2015.
  32. Hochreiter S, Heusel M, Obermayer K. Fast model-based protein homology detection without alignment. *Bioinformatics* 2007;23(14):1728-36.
  33. Sønderby SK, Sønderby CK, Nielsen H et al. Convolutional LSTM Networks for Subcellular Localization of Proteins. *arXiv preprint arXiv:1503.01919* 2015.
  34. Fakoor R, Ladhak F, Nazi A et al. Using deep learning to enhance cancer diagnosis and classification. In: *Proceedings of the International Conference on Machine Learning*. 2013.
  35. Edelman RR, Warach S. Magnetic resonance imaging. *New England Journal of Medicine* 1993;328(10):708-16.
  36. Ogawa S, Lee T-M, Kay AR et al. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences* 1990;87(24):9868-72.
  37. Hsieh J. Computed tomography: principles, design, artifacts, and recent advances. SPIE Bellingham, WA, 2009.
  38. Chapman D, Thominson W, Johnston R et al. Diffraction enhanced x-ray imaging. *Physics in medicine and biology* 1997;42(11):2015.
  39. Bailey DL, Townsend DW, Valk PE et al. Positron emission tomography. Springer, 2005.
  40. Gurcan MN, Boucheron LE, Can A et al. Histopathological image analysis: a review. *Biomedical Engineering, IEEE Reviews in* 2009;2:147-71.
  41. Plis SM, Hjelm DR, Salakhutdinov R et al. Deep learning for neuroimaging: a validation study. *Frontiers in neuroscience* 2014;8.
  42. Hua K-L, Hsu C-H, Hidayati SC et al. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *Oncotargets and therapy* 2015;8.
  43. Suk H-I, Shen D. Deep learning-based feature representation for AD/MCI classification. *Medical Image Computing and Computer-Assisted Intervention–MICCAI* 2013. Springer, 2013, 583-90.
  44. Roth HR, Lu L, Liu J et al. Improving Computer-aided Detection using Convolutional Neural Networks and Random View Aggregation. *arXiv preprint arXiv:1505.03046* 2015.
  45. Roth HR, Yao J, Lu L et al. Detection of sclerotic spine metastases via random aggregation of deep convolutional neural network classifications. *Recent Advances in Computational Methods*

- and Clinical Applications for Spine Imaging. Springer, 2015, 3-12.
46. Li Q, Cai W, Wang X et al. Medical image classification with convolutional neural network. In: *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*. 2014. p. 844-8. IEEE.
47. Cireşan DC, Giusti A, Gambardella LM et al. Mitosis detection in breast cancer histology images with deep neural networks. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer, 2013, 411-8.
48. Ypsilantis P-P, Siddique M, Sohn H-M et al. Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks. *PLoS one* 2015;10(9):e0137036.
49. Zeng T, Li R, Mukkamala R et al. Deep convolutional neural networks for annotating gene expression patterns in the mouse brain. *BMC bioinformatics* 2015;16(1):1-10.
50. Cruz-Roa AA, Ovalle JEA, Madabhushi A et al. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer, 2013, 403-10.
51. Bar Y, Diamant I, Wolf L et al. Deep learning with non-medical training used for chest pathology identification. In: *SPIE Medical Imaging*. 2015. p. 94140V-V-7. International Society for Optics and Photonics.
52. Li Q, Feng B, Xie L et al. A Cross-modality Learning Approach for Vessel Segmentation in Retinal Images. *IEEE Transactions on Medical Imaging* 2015;35(1):109 - 8.
53. Ning F, Delhomme D, LeCun Y et al. Toward automatic phenotyping of developing embryos from videos. *Image Processing, IEEE Transactions on* 2005;14(9):1360-71.
54. Turaga SC, Murray JF, Jain V et al. Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural Computation* 2010;22(2):511-38.
55. Helmstaedter M, Briggman KL, Turaga SC et al. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* 2013;500(7461):168-74.
56. Ciresan D, Giusti A, Gambardella LM et al. Deep neural networks segment neuronal membranes in electron microscopy images. In: *Advances in neural information processing systems*. 2012. p. 2843-51.
57. Prasoon A, Petersen K, Igel C et al. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer, 2013, 246-53.
58. Hawei M, Davy A, Warde-Farley D et al. Brain Tumor Segmentation with Deep Neural Networks. *arXiv preprint arXiv:1505.03540* 2015.
59. Roth HR, Lu L, Farag A et al. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015. Springer, 2015, 556-64.
60. Stollenga MF, Byeon W, Liwicki M et al. Parallel Multi-Dimensional LSTM, With Application

- to Fast Biomedical Volumetric Image Segmentation. *arXiv preprint arXiv:1506.07452* 2015.
61. Xu J, Xiang L, Liu Q et al. Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology images. *IEEE Transactions on Medical Imaging* 2015;35(1):119 - 30.
62. Cho J, Lee K, Shin E et al. Medical Image Deep Learning with Hospital PACS Dataset. *arXiv preprint arXiv:1511.06348* 2015.
63. Lee S, Choi M, Choi H-s et al. FingerNet: Deep learning-based robust finger joint detection from radiographs. In: *Biomedical Circuits and Systems Conference (BioCAS), 2015 IEEE*. 2015. p. 1-4. IEEE.
64. Roth HR, Lee CT, Shin H-C et al. Anatomy-specific classification of medical images using deep convolutional nets. *arXiv preprint arXiv:1504.04003* 2015.
65. Roth HR, Lu L, Seff A et al. A new 2.5 D representation for lymph node detection using random sets of deep convolutional neural network observations. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*. Springer, 2014, 520-7.
66. Gerven MAV, De Lange FP, Heskes T. Neural decoding with hierarchical generative models. *Neural Computation* 2010;22(12):3127-42.
67. Koyamada S, Shikauchi Y, Nakae K et al. Deep learning of fMRI big data: a novel approach to subject-transfer decoding. *arXiv preprint arXiv:1502.00093* 2015.
68. Niedermeyer E, da Silva FL. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.
69. Buzsáki G, Anastassiou CA, Koch C. The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nature reviews neuroscience* 2012;13(6):407-20.
70. Marriott HJL, Wagner GS. *Practical electrocardiography*. Williams & Wilkins Baltimore, 1988.
71. De Luca CJ. The use of surface electromyography in biomechanics. *Journal of applied biomechanics* 1997;13:135-63.
72. Young LR, Sheena D. Eye-movement measurement techniques. *American Psychologist* 1975;30(3):315.
73. Barea R, Boquete L, Mazo M et al. System for assisted mobility using eye movements based on electrooculography. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 2002;10(4):209-18.
74. Freudenburg ZV, Ramsey NF, Wronkeiwicz M et al. Real-time naive learning of neural correlates in ECoG Electrophysiology. *Int. Journal of Machine Learning and Computing* 2011.
75. An X, Kuang D, Guo X et al. A Deep Learning Method for Classification of EEG Data Based on Motor Imagery. *Intelligent Computing in Bioinformatics*. Springer, 2014, 203-10.
76. Li K, Li X, Zhang Y et al. Affective state recognition from EEG with deep belief networks. In: *Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference on*. 2013. p. 305-10. IEEE.
77. Jia X, Li K, Li X et al. A Novel Semi-Supervised Deep Learning Framework for Affective State Recognition on EEG Signals. In: *Bioinformatics and Bioengineering (BIBE), 2014 IEEE International Conference on*. 2014. p. 30-7. IEEE.
78. Zheng W-L, Guo H-T, Lu B-L. Revealing critical channels and frequency bands for emotion

- recognition from EEG with deep belief network. In: *Neural Engineering (NER), 2015 7th International IEEE/EMBS Conference on*. 2015. p. 154-7. IEEE.
79. Jirayucharoensak S, Pan-Ngum S, Israsena P. EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation. *The Scientific World Journal* 2014;2014.
80. Stober S, Cameron DJ, Grahn JA. Classifying EEG recordings of rhythm perception. In: *15th International Society for Music Information Retrieval Conference (ISMIR'14)*. 2014. p. 649-54.
81. Stober S, Cameron DJ, Grahn JA. Using Convolutional Neural Networks to Recognize Rhythm. In: *Advances in Neural Information Processing Systems*. 2014. p. 1449-57.
82. Cecotti H, Graeser A. Convolutional neural network with embedded Fourier transform for EEG classification. In: *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. 2008. p. 1-4. IEEE.
83. Cecotti H, Gräser A. Convolutional neural networks for P300 detection with application to brain-computer interfaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 2011;33(3):433-45.
84. Soleymani M, Asghari-Esfeden S, Pantic M et al. Continuous emotion detection using EEG signals and facial expressions. In: *Multimedia and Expo (ICME), 2014 IEEE International Conference on*. 2014. p. 1-6. IEEE.
85. Wang Z, Lyu S, Schalk G et al. Deep feature learning using target priors with applications in ECoG signal decoding for BCI. In: *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*. 2013. p. 1785-91. AAAI Press.
86. Stober S, Sternin A, Owen AM et al. Deep Feature Learning for EEG Recordings. *arXiv preprint arXiv:1511.04306* 2015.
87. Huanhuan M, Yue Z. Classification of Electrocardiogram Signals with Deep Belief Networks. In: *Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on*. 2014. p. 7-12. IEEE.
88. Wulsin D, Gupta J, Mani R et al. Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement. *Journal of neural engineering* 2011;8(3):036015.
89. Turner J, Page A, Mohsenin T et al. Deep belief networks used on high resolution multichannel electroencephalography data for seizure detection. In: *2014 AAAI Spring Symposium Series*. 2014.
90. Zhao Y, He L. Deep Learning in the EEG Diagnosis of Alzheimer's Disease. In: *Computer Vision-ACCV 2014 Workshops*. 2014. p. 340-53. Springer.
91. Längkvist M, Karlsson L, Loutfi A. Sleep stage classification using unsupervised feature learning. *Advances in Artificial Neural Systems* 2012;2012:5.
92. Mirowski P, Madhavan D, LeCun Y et al. Classification of patterns of EEG synchronization for seizure prediction. *Clinical neurophysiology* 2009;120(11):1927-40.
93. Petrosian A, Prokhorov D, Homan R et al. Recurrent neural network based prediction of

- epileptic seizures in intra-and extracranial EEG. *Neurocomputing* 2000;30(1):201-18.
94. Davidson PR, Jones RD, Peiris MT. EEG-based lapse detection with high temporal resolution. *Biomedical Engineering, IEEE Transactions on* 2007;54(5):832-9.
95. LeCun Y, Ranzato M. Deep learning tutorial. In: *Tutorials in International Conference on Machine Learning (ICML'13)*. 2013. Citeseer.
96. Farabet C, Couprie C, Najman L et al. Learning hierarchical features for scene labeling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 2013;35(8):1915-29.
97. Szegedy C, Liu W, Jia Y et al. Going deeper with convolutions. *arXiv preprint arXiv:1409.4842* 2014.
98. Tompson JJ, Jain A, LeCun Y et al. Joint training of a convolutional network and a graphical model for human pose estimation. In: *Advances in Neural Information Processing Systems*. 2014. p. 1799-807.
99. Liu N, Han J, Zhang D et al. Predicting eye fixations using convolutional neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015. p. 362-70.
100. Hinton G, Deng L, Yu D et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Processing Magazine, IEEE* 2012;29(6):82-97.
101. Sainath TN, Mohamed A-r, Kingsbury B et al. Deep convolutional neural networks for LVCSR. In: *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. 2013. p. 8614-8. IEEE.
102. Chorowski JK, Bahdanau D, Serdyuk D et al. Attention-based models for speech recognition. In: *Advances in Neural Information Processing Systems*. 2015. p. 577-85.
103. Kiros R, Zhu Y, Salakhutdinov RR et al. Skip-thought vectors. In: *Advances in Neural Information Processing Systems*. 2015. p. 3276-84.
104. Li J, Luong M-T, Jurafsky D. A hierarchical neural autoencoder for paragraphs and documents. *arXiv preprint arXiv:1506.01057* 2015.
105. Luong M-T, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025* 2015.
106. Svozil D, Kvasnicka V, Pospichal J. Introduction to multi-layer feed-forward neural networks. *Chemometrics and intelligent laboratory systems* 1997;39(1):43-62.
107. Vincent P, Larochelle H, Bengio Y et al. Extracting and composing robust features with denoising autoencoders. In: *Proceedings of the 25th international conference on Machine learning*. 2008. p. 1096-103. ACM.
108. Vincent P, Larochelle H, Lajoie I et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *The Journal of Machine Learning Research* 2010;11:3371-408.
109. Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Computation* 2006;18(7):1527-54.

110. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science* 2006;313(5786):504-7.
111. Le Cun BB, Denker JS, Henderson D et al. Handwritten digit recognition with a back-propagation network. In: *Advances in neural information processing systems*. 1990. Citeseer.
112. Lawrence S, Giles CL, Tsoi AC et al. Face recognition: A convolutional neural-network approach. *Neural Networks, IEEE Transactions on* 1997;8(1):98-113.
113. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. 2012. p. 1097-105.
114. Williams RJ, Zipser D. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation* 1989;1(2):270-80.
115. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. *Neural Networks, IEEE Transactions on* 1994;5(2):157-66.
116. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation* 1997;9(8):1735-80.
117. Gers FA, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM. *Neural Computation* 2000;12(10):2451-71.
118. Cho K, Van Merriënboer B, Gulcehre C et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* 2014.
119. Graves A, Schmidhuber J. Offline handwriting recognition with multidimensional recurrent neural networks. In: *Advances in neural information processing systems*. 2009. p. 545-52.
120. Hadsell R, Sermanet P, Ben J et al. Learning long-range vision for autonomous off-road driving. *Journal of Field Robotics* 2009;26(2):120-44.
121. Masci J, Meier U, Cireşan D et al. Stacked convolutional auto-encoders for hierarchical feature extraction. *Artificial Neural Networks and Machine Learning–ICANN 2011*. Springer, 2011, 52-9.
122. Minsky M, Papert S. Perceptron: an introduction to computational geometry. *The MIT Press, Cambridge, expanded edition* 1969;19(88):2.
123. Fukushima K. Cognitron: A self-organizing multilayered neural network. *Biological cybernetics* 1975;20(3-4):121-36.
124. Hinton GE, Sejnowski TJ. Learning and relearning in Boltzmann machines. *Parallel distributed processing: Explorations in the microstructure of cognition* 1986;1:282-317.
125. Hinton G. A practical guide to training restricted Boltzmann machines. *Momentum* 2010;9(1):926.
126. Jia Y. Caffe: An open source convolutional architecture for fast feature embedding. In: *ACM International Conference on Multimedia*. ACM. 2014.
127. Palm RB. Prediction as a candidate for learning deep hierarchical models of data. *Technical University of Denmark* 2012.
128. Collobert R, Kavukcuoglu K, Farabet C. Torch7: A matlab-like environment for machine learning. In: *BigLearn, NIPS Workshop*. 2011.

129. Bergstra J, Breuleux O, Bastien F et al. Theano: a CPU and GPU math expression compiler. In: *Proceedings of the Python for scientific computing conference (SciPy)*. 2010. p. 3. Austin, TX.
130. Bastien F, Lamblin P, Pascanu R et al. Theano: new features and speed improvements. *arXiv preprint arXiv:1211.5590* 2012.
131. Chollet F. Keras: Theano-based Deep Learning library. *Code: https://github.com/fchollet. Documentation: http://keras.io* 2015.
132. Goodfellow IJ, Warde-Farley D, Lamblin P et al. PyLearn2: a machine learning research library. *arXiv preprint arXiv:1308.4214* 2013.
133. Abadi M, Agarwal A, Barham P et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. *Software available from tensorflow.org* 2015.
134. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 2010. p. 807-14.
135. Hecht-Nielsen R. Theory of the backpropagation neural network. In: *Neural Networks, 1989. IJCNN, International Joint Conference on*. 1989. p. 593-605. IEEE.
136. Bottou L. Stochastic gradient learning in neural networks. *Proceedings of Neuro-Nimes* 1991;91(8).
137. Erhan D, Bengio Y, Courville A et al. Why does unsupervised pre-training help deep learning? *The Journal of Machine Learning Research* 2010;11:625-60.
138. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature* 2010;463(7280):457-63.
139. Jolliffe I. Principal component analysis. Wiley Online Library, 2002.
140. Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology* 1968;195(1):215-43.
141. Castro-Orós D, Pampín S, Bolado-Carrancio A et al. Functional analysis of LDLR promoter and 5' UTR mutations in subjects with clinical diagnosis of familial hypercholesterolemia. *Human mutation* 2011;32(8):868-72.
142. Stenson PD, Mort M, Ball EV et al. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Human genetics* 2014;133(1):1-9.
143. Duryea J, Jiang Y, Countryman P et al. Automated algorithm for the identification of joint space and phalanx margin locations on digitized hand radiographs. *Medical Physics* 1999;26(3):453-61.
144. Schuster M, Paliwal KK. Bidirectional recurrent neural networks. *Signal Processing, IEEE Transactions on* 1997;45(11):2673-81.
145. Cenic A, Nabavi DG, Craen RA et al. Dynamic CT measurement of cerebral blood flow: a validation study. *American Journal of Neuroradiology* 1999;20(1):63-73.
146. Tsao J, Boesiger P, Pruessmann KP. k-t BLAST and k-t SENSE: Dynamic MRI with high frame rate exploiting spatiotemporal correlations. *Magnetic Resonance in Medicine* 2003;50(5):1031-42.
147. Cohen AM, Hersh WR. A survey of current work in biomedical text mining. *Briefings in*

*bioinformatics* 2005;6(1):57-71.

148. Graves A, Wayne G, Danihelka I. Neural turing machines. *arXiv preprint arXiv:1410.5401* 2014.
149. Rasmus A, Berglund M, Honkala M et al. Semi-Supervised Learning with Ladder Networks. In: *Advances in Neural Information Processing Systems*. 2015. p. 3532-40.
150. Bengio Y. Learning deep architectures for AI. *Foundations and trends® in Machine Learning* 2009;2(1):1-127.
151. He H, Garcia EA. Learning from imbalanced data. *Knowledge and Data Engineering, IEEE Transactions on* 2009;21(9):1263-84.
152. López V, Fernández A, García S et al. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences* 2013;250:113-41.
153. Oh S, Lee MS, Zhang B-T. Ensemble learning with active example selection for imbalanced biomedical data classification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 2011;8(2):316-25.
154. Malin BA, El Emam K, O'Keefe CM. Biomedical data privacy: problems, perspectives, and recent advances. *Journal of the American medical informatics association* 2013;20(1):2-6.
155. Sonnenburg S, Schweikert G, Philips P et al. Accurate splice site prediction using support vector machines. *BMC bioinformatics* 2007;8(Suppl 10):S7.
156. Pan SJ, Yang Q. A survey on transfer learning. *Knowledge and Data Engineering, IEEE Transactions on* 2010;22(10):1345-59.
157. Deng J, Dong W, Socher R et al. Imagenet: A large-scale hierarchical image database. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. 2009. p. 248-55. IEEE.
158. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. *Computer Vision–ECCV 2014*. Springer, 2014, 818-33.
159. Karpathy A, Johnson J, Li F-F. Visualizing and understanding recurrent networks. *arXiv preprint arXiv:1506.02078* 2015.
160. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* 2014.
161. Xu K, Ba J, Kiros R et al. Show, attend and tell: Neural image caption generation with visual attention. *arXiv preprint arXiv:1502.03044* 2015.
162. Choromanska A, Henaff M, Mathieu M et al. The loss surfaces of multilayer networks. *arXiv preprint arXiv:1412.0233* 2014.
163. Dauphin YN, Pascanu R, Gulcehre C et al. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In: *Advances in neural information processing systems*. 2014. p. 2933-41.
164. Goodfellow I, Bengio Y, Courville A. Deep Learning. 2016.
165. Bengio Y. Practical recommendations for gradient-based training of deep architectures.

- Neural Networks: Tricks of the Trade. Springer, 2012, 437-78.
166. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *The Journal of Machine Learning Research* 2012;13(1):281-305.
167. Bergstra JS, Bardenet R, Bengio Y et al. Algorithms for hyper-parameter optimization. In: *Advances in Neural Information Processing Systems*. 2011. p. 2546-54.
168. Ngiam J, Khosla A, Kim M et al. Multimodal deep learning. In: *Proceedings of the 28th international conference on machine learning (ICML-11)*. 2011. p. 689-96.
169. Cao Y, Steffey S, He J et al. Medical Image Retrieval: A Multimodal Approach. *Cancer informatics* 2014;13(Suppl 3):125.
170. Ngiam J, Coates A, Lahiri A et al. On optimization methods for deep learning. In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 2011. p. 265-72.
171. Sutskever I, Martens J, Dahl G et al. On the importance of initialization and momentum in deep learning. In: *Proceedings of the 30th international conference on machine learning (ICML-13)*. 2013. p. 1139-47.
172. Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. *The Journal of Machine Learning Research* 2011;12:2121-59.
173. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* 2015.
174. Martens J. Deep learning via Hessian-free optimization. In: *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 2010. p. 735-42.
175. Raina R, Madhavan A, Ng AY. Large-scale deep unsupervised learning using graphics processors. In: *Proceedings of the 26th annual international conference on machine learning*. 2009. p. 873-80. ACM.
176. Ho Q, Cipar J, Cui H et al. More effective distributed ml via a stale synchronous parallel parameter server. In: *Advances in neural information processing systems*. 2013. p. 1223-31.
177. Bengio Y, Schwenk H, Senécal J-S et al. Neural probabilistic language models. *Innovations in Machine Learning*. Springer, 2006, 137-86.
178. Li M, Andersen DG, Park JW et al. Scaling distributed machine learning with the parameter server. In: *11th USENIX Symposium on Operating Systems Design and Implementation (OSDI 14)*. 2014. p. 583-98.
179. Dean J, Corrado G, Monga R et al. Large scale distributed deep networks. In: *Advances in Neural Information Processing Systems*. 2012. p. 1223-31.
180. Kin H, Park J, Jang J et al. DeepSpark: Spark-Based Deep Learning Supporting Asynchronous Updates and Caffe Compatibility. *arXiv preprint arXiv:1602.08191* 2016.
181. Simonite T. Thinking in Silicon. *MIT Technology Review*, 2013.
182. Ovtcharov K, Ruwase O, Kim J-Y et al. Accelerating deep convolutional neural networks using specialized hardware. *Microsoft Research Whitepaper* 2015;2.
183. Farabet C, Poulet C, Han JY et al. Cnp: An fpga-based processor for convolutional networks. In: *Field Programmable Logic and Applications, 2009. FPL 2009. International Conference on*. 2009.

p. 32-7. IEEE.

184. Hof RD. Neuromorphic Chips. MIT Technology Review, 2014.