Deep learning aiding Health Informatics in Drug discovery

Neha Sharma
Mtech Scholar
Sri Sai College of engineering and technology

Deepak Kumar
Assistant Professor
Sri Sai College of engineering and technology

ABSTRACT

The changes that occur are exciting and more challenging in our industry. There has been a massive increase in the amount of data in health informatics in the last decade. Over the years, deep learning is raising with its extraordinary success in the research areas of artificial intelligence. If we form larger neural network and then we train it with more available data and fast enough computers their performance then continues to arise. Earlier, machine-learning tools such as QSAR (quantitative structure-activity relationship) modeling have been used to identify potential biological active molecules from millions candidate compounds for drug discovery. But, in this era of big data machine learning approaches lack efficiency. Hence, deep learning evolved as a solution to the problem of big data. In this paper we discuss various deep learning approaches studied for various applications of health informatics with special reference to drug discovery.

KEYWORDS

deep learning, deep neural network, convolutional neural network, recurrent neural network, deep autoencoder, health informatics

1 Introduction

The volume of digital data in the world is increasing epidemically. In fact, 90% of the world's existing data has been created in the last two years - a dozen or so 2.5 billion bytes of subsidiary data have been created [1]. Machine learning algorithms such as deep learning have been used due to increasing data and strong demand for control. Number of applications such as colorization of black and white images, generation of image captions, character text generation, automatic play, automatic translation, natural language processing, Bio informatics etc. [2]. As a result of these advances, the monopoly of life sciences has experienced an unplanned growth of the bulk of achievable composite activities and biomedical data to bring out trends and technologies. As the data is large scale, this requires its excavation for drug discovery. Machine learning methods such as neural networks, support vector machines, and random forests were used to improve the QSAR model, deep learning, and matrix factorization. There is a steady increase in the power of the computer and the amount of data that is conducive to deep learning [3]. Deep learning (DL) is better than other machine learning methods because of the flexible nature of neural network architecture in deep approaches. This review includes convolutional neural networks (CNNs), recurrent neural networks (RNNs), and fully connected wireless networks. For a long time, single layer neural networks have been used in QSAR modeling. In addition, data size and estimation power have increased, making multilayer networks unstable for divination of bioactivity.

1.1 Axioms of deep learning

Deep learning is about how we learn and process information. Every action we do or the memory we have is controlled by our nervous system. We use Artificial neural networks to learn the representation of the data [5]. In Artificial neural network we have one well known technique- Deep neural network which consist of no. of layers between the input layer and the output layer with a certain level of complexity. It uses mathematical modeling for the processing of the data in complex ways. The basic structure of modern artificial neural networks is illustrated in Figure 1. When the network moves across the layers it calculates the output. Basically, there are 3 layers in an ANN: the input layer, the hidden layer and the output layer. The nodes that are present in the network are either fully or partially connected and are also called neurons. The nodes have values from 0.0 to 1.0 where 1 represents the value "fully active" and 0 completely deactivates with the other values between 0 and 1. In the network, the nodes are linked by connections where each connection to a "weight". represented by "w". These signal values then procreate from the finished inputs the connection weights "w" to the hidden nodes and then pass through more connection weights to the output nodes [6]. The probability between the input and output values of a hidden unit can be symbolized in Figure 2.

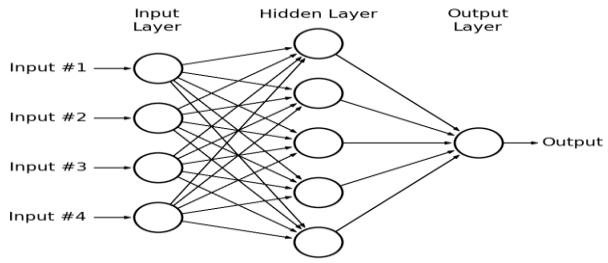

Figure 1: Artificial neural network.

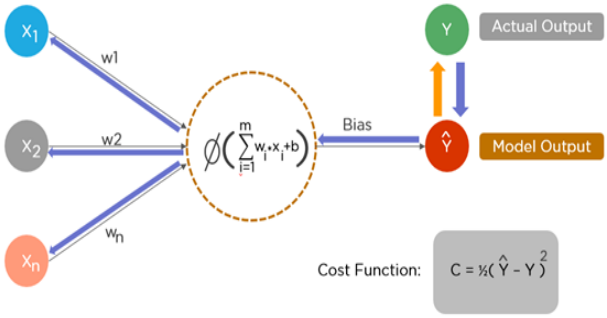

Figure 2: Calculation of output values from input values using cost function

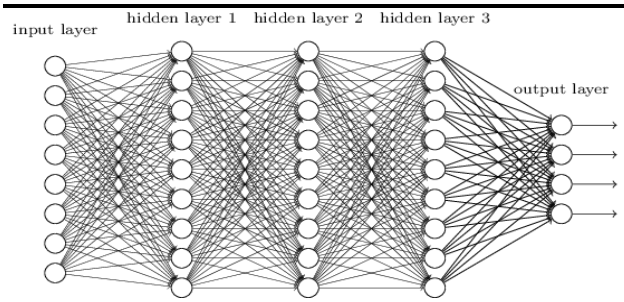

Figure 3: Deep learning using large number of hidden layers

Back propagation method is followed to refine the errors between the true and predicted value. In the learning process of ANN, the weight values in the network are adjusted imperatively. In the network, the gradient which is required for the calculation of the weights to be used in the network is calculated by the back propagation. In the learning environment, the gradient descent optimization algorithm uses backward propagation by computing the gradient of the loss function to adjust the weight of the neurons. This technique is also known as inverse error propagation because the error is calculated at the output and, if necessary, is redistributed across the network layers. Using a cost function, the error in the network is calculated [7]. The neural network implements any type of deep learning algorithm such as CNN, RNN that will be discussed in this review. The modern structure of the artificial neural network algorithm was developed in the 1960s to 1980s and the number of applications appeared while the traditional structure of artificial neural networks encountered problems such as overfitting etc was predominantly regained by machine learning algorithm such as support vector machine [8] and RF [9]. There are many differences between deep learning and the traditional artificial neural network. Traditional ANNs can result in one or two hidden layers, while deep learning uses a large number of hidden layers, as shown in Figure 3. In case of deep learning no. nodes can be used in each layer due to the appearance of more efficient CPU and GPU hardware. In deep learning methods and DropConnect to solve problems, a linear rectified and adjusted unit is applied to avoid leakage gradients. Most of the time, DL packages are open-source [10]. The NN architectures used in deep learning are DNN, CNN, and RNN. DNN stands for deep neural network (Figure 3), which is an artificial neural network which is consist of multiple layers between input and output layers. DNNs may include a non-linear mosaic relationship. As data do not travel back in DNN and travels from input to the output layers, so these networks are considered as the safe networks. DNN can retrieve entities at different hierarchical levels [11]. Second technique under artificial neural network is the convolutional neural network (CNN or ConvNet) (Figure 4). These use small scale pre-processing compared to all other preprocessing algorithms. CNNs are mainly used to analyze visual images. The structure of CNN consists of several layers of convolution and subsampling. The convolutional layers apply the convolution operation to the input. Each convolutional neuron tracks data only for its receptive field. The convolutional layer in CNN consists of kernels with learnable parameters and a small receptive field where the kernel is convulsed over the height and width of the input volume, then calculates the dot product between the kernel inputs produces a map. features that is 2D of this kernel. The size of the feature map decreases with depth, CNN layers near the input layer will have fewer filters, while layers that are far apart will have more filters in comparison. If we want to equalize the computation at each layer, the product of the position x pixel of the function must be kept almost constant on all the layers. The number of feature maps depends on the number of available examples and the complexity level of the task, as well as the capacity. The common filter forms are chosen based on the dataset and the challenge is to find the right level of granularity in order to create an appropriate abstraction for the chosen or given dataset. There are three hyperparameters: the deep, stride and zero fill that controls the size of the output volume of the convolutional layer. The number of neurons in the layer that connects to the same region of the input volume is controlled by the depth of the output volume. The allocation of depth columns around spatial dimensions is controlled by the stride. One pixel at a time is moved when the stride is "1" and when it is "2", two pixels at a time are moved. The size of the padding is a third parameter, it controls the spatial size of the output volume. Finally, the feature maps are concatenated into fully connected layers where the neighboring neurons of the network are connected like the traditional ANNs and give the final output value. CNN reduces memory consumption and increases learning speed [12].

There is a variant of an artificial neural network: RNN (Figure 5). In this network, there is a connection between the nodes that form a directed graph with the sequence in a successive layer format. These networks use their memory for input processing in order, making it relevant for tasks such as voice recognition, handwriting recognition, and so on [13]. LSTM combined with CNN has improved in the field of automatic captioning of images. The problem of the disappearance gradient is reduced by this network [14]. The network of Figure 6 is called auto-encoder (AE), used to learn an illustration for a set of data to reduce dimensionality. It compresses the data from the input layer to the abbreviated code and then decompresses the code into something that matches the data that is original [15]. Automatic encoders can be stacked on top of each other to create a deep automatic encoder architecture. It is also confronted with the problem of the disappearance of gradients during training. Many variations have been proposed for the learned representations to be robust or stable in the case of small variations in the pattern of entry. Finally, convolutional auto coding was the solution to this problem, which dealt with two-dimensional images and effectively preserved the spatial locality [16].

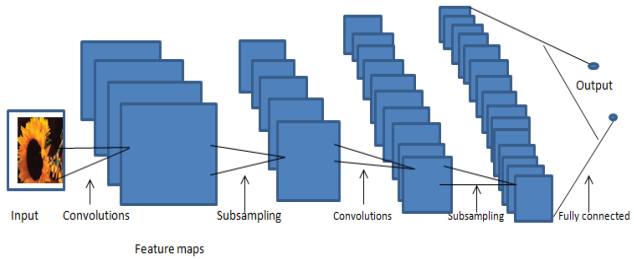

Figure 4: Convolutional neural network

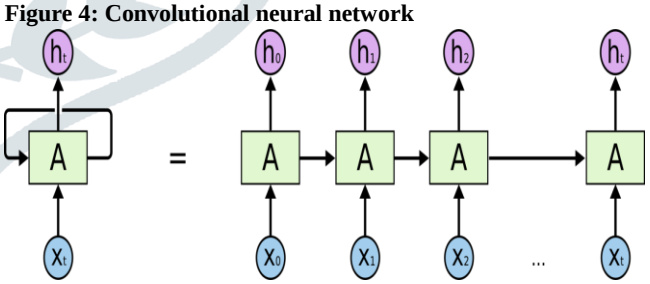

Figure 5: Recurrent neural network

Table 1 Different deep learning architectures

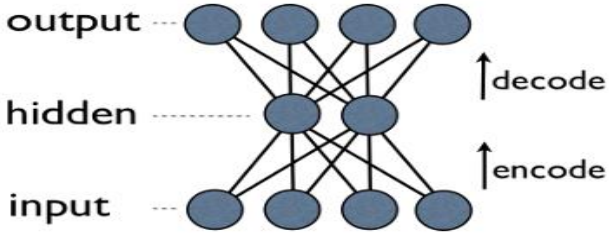

Figure 6: Auto-encoder

ARCHITECTURE	DESCRIPTION	PROS	CONS	PUBLICATION
Deep neural network	Neural network with multiple hidden layers which allows non-linear hypotheses	Handles unstructured data Successful in many applications	Learning process can be slow	[22-25]
Convolutional neural network	Highly accurate for 2D mages Inspired by biological processes	Few neuron connections required as compared to typical NN Applicable in image and video recognition, recommender systems	Large number of layers required Large dataset required	[14-18]
Recurrent neural network	Sequenced directed graph is formed by connection of nodes Same weight is shared across all steps in the network Examines stream of data	Uses internal memory to process sequential events Dynamic temporal behavior for time sequences Applicable in handwriting and speech recognition	Vanishing gradient problem causes learning issues	[13]
Deep autoencoder	Feature extraction or dimentionality reduction is the main objective Learns efficient data codings in an unsupervised learning method Recently used for learning of generative models of data Has a number of input and output nodes	Labeled data is not required for training Number of techniques exist that improve the ability to capture important information Representation with higher level are stable and robust	Results in very slow learning process as well as poor solutions which can be solved by using initial weights which is referred to as pre-training Errors when back-propagated become insignificant	[5-9]

2 Application areas of Health Informatics

Health informatics is information engineering applied to the field of health care, generally the management and use of information relating to the care of patients. It is a multidisciplinary field that involves disciplines such as information science, computer science, social sciences, behavioral sciences, management sciences and others. It applies to the areas of nursing, clinical medicine, dentistry, pharmacy, public health, occupational therapy, physiotherapy, etc. The massive growth and widespread availability of digital health data over the past decade has led to increased research activity in the health care field. Traditional approaches to health data management have had limited success because they are unable to handle the large amount of complex data with high volume, high speed, and variety. The following table describes the use of deep learning in various health informatics applications.

Table 2 Applications of health informatics

AREA	APPLICATIONS	DATA SOURCE	TECHNIQUE	REFERENCE
Bioinformatics	Cancer diagnosis	Gene expression	Deep autoencoder	[34]
	Gene variants	Microarray data	Deep neural networks	[35]
	DNA methylation	Genes/RNA/DNA sequences	Deep neural networks	[36]
	Drug design	Molecule compounds	Deep neural networks	[37]
Medical imaging	Neural cell classification	MRI/fMRI PET scans	Convolutional neural networks	[38-41]
	Tumor detection	X-ray images	Deep neural network	[42-43]
Pervasive sensing	Human activity recognition	Video wearable device	Convolutional neural networks	[44-47]
	Food intake	wearable device	Deep neural network	[48]
	Energy expenditure	RGB image Mobile device	Convolutional neural networks	[49]
			Deep neural network	[50]
Medical informatics	Prediction of disease	Electronic health records	Convolutional neural networks	[51]
			Deep autoencoder	[52-53]
	Data mining	Blood/lab tests	Deep neural network	[54-55]
			Recurrent neural networks	[52, 56]
Public health	Predicting demographic information	Social media data	Deep autoencoder	[57]
	Air pollutant prediction	Text messages	Deep neural network	[58-61]

3 Literature Survey

With a massive influx of multimodal data, the role of data analysis in health informatics has grown rapidly over the last decade. This has also generated a growing interest in the production of data-driven analytical models based on machine learning in health informatics. Deep learning, a technique based on artificial neural networks, is becoming in recent years a powerful tool for machine learning, promising to reshape the future of artificial intelligence. The theoretical underpinnings of deep learning are well anchored in the literature on classical neural networks. But, unlike the more traditional use of NNs, deep learning explains the use of many hidden neurons and layers, usually more than two, as an architectural benefit associated with new training paradigms. It provides a high level of abstraction that makes the feature sets automatic. In areas such as health informatics, the generation of this automatic set of features without human intervention has many advantages. For example, in medical imaging, it can generate more sophisticated features that are difficult to develop by descriptive means. Neural convolution networks (CNN) have had the greatest impact in the field of health informatics. Other plausible architectures for deep learning include those based on Restricted Boltzmann (RBM) machine compositions such as Deep Belief Networks (DBNs), stacked auto-encoders that function as deep autoencoders. Table 2 presents various health informatics studies that have used deep learning in health informatics.

Deep CNNs are the subclass of the deep neural network that have been used in drug discovery for the prediction of ligand-based bioactivity and QSAR. AtomNet is the first CNN system-based system designed to predict the bioactivity of small molecules for drug discovery applications. The functioning of biological systems can only occur when there is a physical interaction between the molecules. It is essential for the discovery of new drugs that, in the case of molecular linkage and contempt, three decades of business computing are still too imprecise for the prediction of routine binding. If the ability to predict

molecular binding will be accurate and efficient, it will reduce the time needed to discover new treatments. In case of development, the toxic molecules can be eliminated in the early stages and can guide the efforts of medicinal chemistry. For speech and vision, the deep convolutional neural networks are best for prediction. It constrains its model architecture to capture the temporal and spatial structure of its domain, reducing the problem of over-adjustment. A deep neural network has been established - RF (random forest) and SVM (carrier vector machine) for virtual screening based on QSAR and ligands. Dahl et al. introduced an architecture for kagglemerk molecular activity challenge that was a multi-tasking deep neural network that forms a unique neural network architecture with a number of output neurons. These multiple output neurons predict the activity of the input molecule in a different evaluation. As the molecules are tested in a number of different estimates, the multi-task deep neural network merges proof of training into almost identical prediction tasks. He used a large number of 2D topological descriptors and DNN showed slightly better performance in 13 of the 15 targets than the standard RF method. Ramsundar et al. methodical study accomplished to build multi-tasking DNN models. His study shows that deep multi-task neural networks work or work better than the single task. Therefore, the multi-tasking DNN is very efficient and gives better results. Koutsoukas et al. studied and compared a deep neural network model with automatic learning methods such as SVM, RF, on seven ChEmBL datasets. Deep neural networks have been found by performing other methods of machine learning. Aliper et al. Constructs a deep neural network for predicting the pharmacological properties of drugs and for drug reprocessing by dragging the LINCS project's metabolomic data as well as lane information. The conclusion is that the DNN model has achieved great accuracy in predicting drug indications. As previously discussed about AtomNet, which is the first structure-based deep convolution neural network designed for the prediction of small molecule bioactivity for drug discovery applications. Atomnet can predict new active molecules for targets without modulators. He showed unexpected results on the structure-based benchmark. In the next section, Table 3 provides a comparative analysis of these studies.

4 Research Gaps

In this era of digital data, health informatics has certainly become an important field of interest. There are countless application areas of health informatics out of which drug discovery was the one under consideration for this review paper. Initially, the study of drug discovery started with machine learning's QSAR model which included SVM and RF. ANN's were sometimes used utilized with QSAR. With increase in data, demand for more effective methods increased. This is when deep learning cane into play. Undoubtedly deep learning has been very much successful in drug discovery but still there are various loopholes. To begin with, there is lack of knowledge and interest among the pharmaceutical and stakeholders in drugs. Adding to that is the lack of knowledge about diseases. And at that, the deep learning approaches CNNs, RNNs, Deep Autoencoders have their own limitations. These architectures are certainly not suitable for smaller datasets. They face a problem while dealing with compounds of similar structure and different activities. Some of these, fail to handle temporal healthcare data. Many other problems can be stated. But, the focus should be laid on generating targeted medicines as bringing medicines into the market is an expensive task. Moreover, combination of deep learning approaches with other machine learning methods should be tried to achieve better performances and better drug discovery.

5 Conclusion and future scope

Deep learning has gained immense popularity in the field of machine learning and pattern recognition. However, deep learning generally requires large sets of data for training purposes. The question is how to learn with the small amount of data available. Medicinal chemists typically work with a limited amount of data at their disposal with new targets, which is relevant to the one-time method of learning drug discovery. Lately, deep learning has used a new type of architecture, namely augmented memory neural networks, the first version being the neural cutting machine that has been significantly improved with a differentiable neural computer (DNC). Then, the DNCs were applied to find the shortest paths and answer the questions. However, these have not yet been used in drug discovery. In-depth learning enables the development of data-driven solutions in health informatics by enabling the automatic generation of entities that reduce the amount of human interference. This has proven advantageous for unstructured data produced by medical imaging, bioinformatics, medical informatics, etc. Nevertheless, deep learning can-not be considered the only solution to increase data in health informatics. But if one wants the studies to be considered, deep learning has proven to be better than any

other method of machine learning. Nevertheless, deep learning cannot be considered the only solution to increasing data in health informatics.

The following table contains the details of various studies carried out in the field of drug discovery using deep learning.

Table 3 Comparison of literature

Author's name	Title Of paper	Journal/ Conference	Year	Description/ Techniques used	Merits	Demerits	Future work
Hongming Chen et al.	The rise of deep learning in drug discovery [8]	Elsevier	2018	DNN CNN RNN AE	Flexible CNNs Successful in computer vision RNNs reduce vanishing gradient problem AEs reduce non-linear dimensionality	Not suitable for small datasets	Dealing small datasets
Chris Molloy	Drug discovery tomorrow: how to catapult ourselves into the future [62]	Elsevier	2018	Real world patient data discovered and mixture of agile companies working together	Provide industrial expertise Prove new technologies Efficient	Bringing medicines into market is costly and time consuming Large scale clinical trial failures	Generate targeted medicines
Youjun Xu et al.	Deep learning for Drug-Induced Liver Injury [65]	ACS AuthorChoice (journal of chemical information and modeling)	2018	DL architecture (DL models, DL-NCTR DILI model, DL-combined DILI model, DL-liew DILI model,)	Performs well in predicting large external validation sets No uncertainty of data DL-combined model suitable for drug DILI risk prediction Good labeling strategy adopted in liew dataset	DILI prediction model less efficient than DL	Covering subjects beyond DILI prediction
Lu Zhang et al.	From machine learning to deep learning: progress in machine intelligence for rational drug discovery [63]	Elsevier	2017	QSAR modeling	Fast Cost effective Deals with high dimensional variables in small datasets DT avoid overfittingusing pruning	Lack of rational interpretation of associated biological mechanism Cannot deal with two compounds with similar structure and different activities	Better performance with deep learning

Daniele Ravi et al.	Deep learning for health information [17]	IEEE	2017	RBM based technique, Deep belief network, Deep Boltzmann, DNN, CNN, RNN AE,	Reduced uncertainty in decision making process	Deep learning can slow down the development of other machine learning techniques	Machine learning approaches other than deep learning can be used
Ricardo miotto et al.	Deep learning for healthcare: review, opportunities and challenges [64]	Oxford	2017	CNN, LSTM-RNN, Stacked denoising AE, Stacked AE, GRU-RNN, Stacked sparse AE , RBM	Healthcare systems can include millions of patient records Applicable in disease risk prediction, personalized prescriptions etc.	Lack of complete knowledge of diseases Difficult to handle temporal healthcare data	Need to estimate network parameters properly Need to consider data sparsity, redundancy and missing values
Han Altae-Tran et al.	Low data drug discovery with one-shot learning [66]	ACS authorchoice (ACS central science)	2017	Mathematical formalism One-shot learning Iterative refinement LSTM Graph convolutions	Simpler methods for low-data learning		Investigate the structure of the embeddings learned by the iterative refinement LTSM modules.
Marcus Olivecrona et al.	Molecular de-novo design through deep reinforcement learning [67]	Journal of cheminformatics	2017	RNN	Promising and robust method for molecular de novo design Data driven, does not rely on pre-existing building blocks	Can't handle multiple parameters such as target activity, DMPK profile, synthetic accessibility etc. simultaneously	Token embeddings. Multi-parametric scoring functions
Izhar Wallach et al.	AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery [68]	arxiv.org	2015	DCNN	First DCNN for molecular binding affinity prediction Incorporates structural information about the target to make its predictions	Not mentioned	Greater AUC on more percentage of targets

Evgeny Byvatov	Comparison of support vector machine and artificial neural network system for drug/nondrug classification [69]	ACS	2003	SVM	Higher overall prediction accuracy of SVM than ANN SVM- useful for classification in QSAR modeling and sparsness of the solution	Not suitable for non-identical sets of correctly and misclassified compounds	Deep learning approaches can be useful
				ANN	Supports large number of input variables		

.

REFERENCES

[1] Jacobson Ralph (2013, April 24) 2.5 quintillion bytes of data created every day. How does CPG & Retail manage it? https://www.ibm.com/blogs/insights-on-business/consumer-products/2-5-quintillion-bytes-of-data-created-every-day-how-does-cpg-retail-manage-it/ (accessed on august 15, 2018)

[2] Goodfellow, I., Bengio, Y., Courville, A., &Bengio, Y. (2016). Deep learning (Vol. 1). Cambridge: MIT press.

[3] Yao, X. J., Panaye, A., Doucet, J. P., Zhang, R. S., Chen, H. F., Liu, M. C., ... & Fan, B. T. (2004). Comparative study of QSAR/QSPR correlations using support vector machines, radial basis function neural networks, and multiple linear regression. Journal of chemical information and computer sciences, 44(4), 1257-1266.

[4] https://en.wikipedia.org/wiki/Deep_learning (accessed on august 15, 2018)

[5] Bilal AfaanArtificial Neural Networks and Deep Learning. Retrieved from https://becominghuman.ai/artificial-neural-networks-and-deep-learning-a3c9136f2137 (accessed on august 15, 2018)

[6] King Paul (2016, April 17) https://www.quora.com/How-do-artificial-neural-networks-work (accessed on august 15, 2018)

[7] Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

[8] Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., &Blaschke, T. (2018). The rise of deep learning in drug discovery. Drug discovery today.

[9] Cortes, C. and Vapnik, V. (1995) Support-vector networks. Mach. Learn. 20, 273–297

[10] Ho, T.K. (1998) The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20, 832–844

[11] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.

[12] Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2011). Unsupervised learning of hierarchical representations with convolutional deep belief networks. Communications of the ACM, 54(10), 95-103.

[13] Fernández, S., Graves, A., &Schmidhuber, J. (2007). An application of recurrent neural networks to discriminative keyword spotting. In International Conference on Artificial Neural Networks (pp. 220-229). Springer, Berlin, Heidelberg.

[14] Sak, H., Senior, A., &Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In Fifteenth annual conference of the international speech communication association.

[15] Bengio, Y. (2009) Learning deep architectures for AI. Found. Trends Mach. Learn. 2, 1–127

[16] Krizhevsky, A., & Hinton, G. E. (2011, April). Using very deep autoencoders for content-based image retrieval. In ESANN.

[17] Ravı, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2017). Deep learning for health informatics. IEEE journal of biomedical and health informatics, 21(1), 4-21.

[18] 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.

[19] 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.

[20] 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.

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

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

[23] 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

[24] 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.

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

[26] 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.

[27] 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.

[28] 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.

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

[30] 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.

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

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

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

[34] 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.

[35] 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.

[36] 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.

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

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

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

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

[41] 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.

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

[43] 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.

[44] 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.

[45] 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.

[46] 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

[47] 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.

[48] 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.

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

[50] P. Kuhad, A. Yassine, and S. Shimohammadi, "Using distance estimationand 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.

[51] 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

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

[53] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep patient: An unsupervised representation to predict the future of patien

[54] 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.

[55] 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.

[56] 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

[57] 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.

[58] 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.

[59] 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

[60] 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.

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

[62] Molloy, C. (2017). Drug discovery tomorrow: how to Catapult ourselves into the future.

[63] Zhang, L., Tan, J., Han, D., & Zhu, H. (2017). From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug discovery today*.

[64] Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2017). Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*.

[65] Xu, Y., Dai, Z., Chen, F., Gao, S., Pei, J., & Lai, L. (2015). Deep learning for drug-induced liver injury. *Journal of chemical information and modeling*, *55*(10), 2085-2093.

[66] Altae-Tran, H., Ramsundar, B., Pappu, A. S., & Pande, V. (2017). Low data drug discovery with one-shot learning. *ACS central science*, *3*(4), 283-293.

[67] Olivecrona, M., Blaschke, T., Engkvist, O., & Chen, H. (2017). Molecular de-novo design through deep reinforcement learning. *Journal of cheminformatics*, *9*(1), 48.

[68] Wallach, I., Dzamba, M., & Heifets, A. (2015). AtomNet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv preprint arXiv:1510.02855*.

[69] Byvatov, E., Fechner, U., Sadowski, J., & Schneider, G. (2003). Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *Journal of chemical information and computer sciences*, *43*(6), 1882-1889.