International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 12 | Dec 2024 www.irjet.net p-ISSN: 2395-0072

Deep Learning in Drug Discovery and Pharmaceutical Research

Kallesh S C[1], Siddesh D S[2], Thejaswi M R[3], Kushi D N[4]

[1]*Bachelor of Engineering, Information Science and Engineering, Bapuji Institute of Engineering and technology, Karnataka, India[1,2]*
[2]*Bachelor of Engineering, Information Science and Engineering, Bapuji Institute of Engineering and technology, Karnataka, India [3,4]*

Abstract - Deep learning has revolutionized the field of drug discovery and pharmaceutical research. This paper provides a comprehensive review of the applications of deep learning in drug discovery and pharmaceutical research. We discuss the various deep learning architectures and techniques used in drug discovery, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs). We also highlight the challenges and limitations of using deep learning in drug discovery and pharmaceutical research.

Key Words: bioinformatics · cheminformatics · drug design · machine-learning · neural network · virtual screening

1.INTRODUCTION

Machine-learning provides a theoretical framework for the discovery and prioritization of bioactive compounds with desired pharmacological effects and their optimization as drug-like leads. Biological target identification and protein design are emerging areas of application. Among the many machine-learning approaches in molecular informatics, chemocentric methods have found widespread application. Their underlying logic typically follows three steps. First, there is the selection of a problem-specific set of descriptors that are believed to capture the essential properties of the molecules involved. At present, there are over 5000 diverse molecular representations ("descriptors") that address the various properties of chemical entities.[1] Second, a metric or scoring scheme is used to compare the encoded molecules to one another.[2] Finally, a machine-learning algorithm is employed to identify the features that may serve to qualitatively or quantitatively distinguish the active from the inactive compounds.[3] Artificial neural networks (ANNs) were among the first methods borrowed from the computer sciences for this purpose.

In 2013, public attention was drawn to a multi-problem QSAR machine-learningchallengein drug discovery posted by Merck. This competition on drug property and activity prediction was won by a deep learning network with a relative accuracy improvement of approximately 14% over Merck's in-house systems and resulted in an article in The New York Times. [5] Here, we present state-of-the-art of advanced chemocentric machine-learning methods with a focus on emerging "deep learning" concepts. We highlight recent advances in the field and point to prospective applications and developments of this potentially game changing technology for drug discovery. Ageneral task for machine-learning is to uncover the relationship between the molecular descriptors used and the measured activity of the compounds to obtain qualitative classifiers or quantitative structure-activity relationship (QSAR) models. Feature extraction from the descriptor patterns is the decisive step in the model development process.[6,7] In current cheminformatics applications, the prevalent machine-learning architectures are "shallow" and contain a single layer of feature transformation. These architectures include linear and nonlinear principle component analysis, k-means clustering methods, partial least square projection to latent structures, decision trees, multivariate linear regression, linear discriminant analysis, support vector machines (SVMs), logistic and kernel regression, multi-layer Perceptrons and related neural network approaches.[8] Although all of these methods have proven to be useful for (Q)SAR modeling and molecular design,[9] the single feature transformation step into a suitable space for the subsequent application of a linear pattern separation model might limit their modeling and representational power when applied to more complex data and setups. A reason for their success in pharmacological applications may stem from the fact that a major part of the complexity inherent to molecular interactions has been engineered into the descriptors employed as patterns for model training, thereby allowing single layer machinelearning architectures to tackle the problem.[10] One challenging question is whether the underlying data complexity and hidden features can be more efficiently dealt with by shifting attention from descriptor engineering to the architecture of the machine-learning system and the training of the algorithms involved. This is the domain of "deep learning".

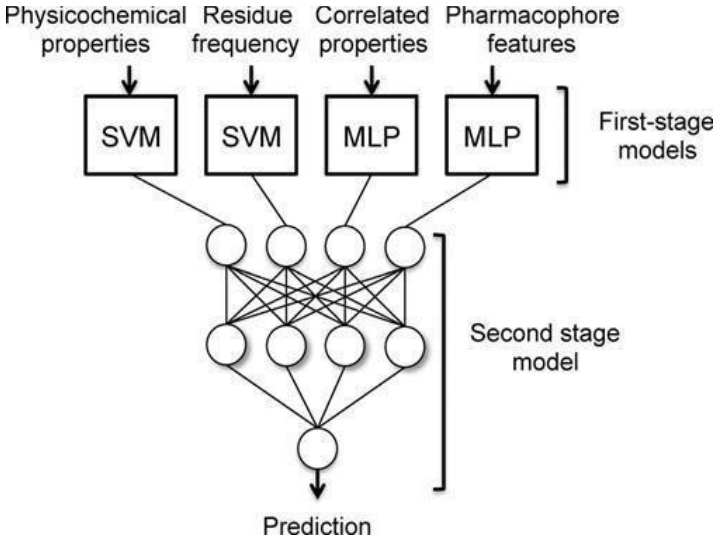

Figure 1. Schematic of a jury network (adapted from ref. 31). In this particular architecture, feature extraction is performed in two stages. The second- stage model is a feed-forward network that weighs the relevance of the first-stage model and improves the overall prediction performance. Note that each first- stage model receives a different set of molecular descriptors as input. SVM, support vector machine; MLP, multi-layer Perceptron.

2 Properties of Deep Learning Architectures

There are several advantages of deep learning neural net- works that suggest that they should replace shallow ANNs:

- A deep, hierarchical architecture enables ANNs to per-form many nonlinear transformations, which leads to the learning of more abstract features compared to shallow networks.[53] This development allows for a more sophisticated combination of low-level features (i.e., nonlinear descriptor combinations) to achieve a better molecular representation and thus classification of compounds. This approach reduces the need for the a priori engineering of more sophisticated descriptors. More abstract nonlinear features tend to be invariant to local changes in the input, leading to inherent noise reduction and increased network robustness.

Therefore, deep learning architectures capture certain families of functions exponentially more efficiently than shallow architectures.

- Deep architectures promote the reuse of features. This ensemble view of the same data corresponds to an improvement on the principle of cascading networks that share the intelligence of different algorithms. How well- regularized deep learning networks can actually exploit commonalities between different tasks to transfer knowledge as inductive bias remains a matter of debate. This aspect is relevant for missing classes within the training set, which is typical for large QSAR data sets. Aside from enabling "educated guesses" on missing information, multitask learning also has the benefits of presenting a shared, learned feature extraction pipeline for multiple tasks. Dahl suggested that this concept might have a regularization effect because weights tend to develop in a way that is useful for many targets instead of overfitting one target in particular.[54] All of these findings promise the simultaneous testing of bioactivities against many related targets at a realistic computational cost. Such applications may be driven even further by the field of polypharmacology.[55] In this case, deep learning might accompany current kernel methods to find additional applications for drugs ("re-purposing") and to assist in the identification of undesired off-target activities.

- As highlighted by the Merck QSAR contest,[5] deep multitask neural networks can be successfully applied to QSAR modeling.[57] Generative deep architectures complement evolutionary algorithms in computational drug discovery that are often used to find new compounds with specific features identified by some previously trained machine-learning architecture.[58] A trained deep generative neural network has learned to separate compound activity classes within its architecture while being able to generate output according to the learned representation; thus, this network contains both functions in one algorithm. Therefore, deep learning might provide a fresh approach to solving the "inverse QSAR problem".

However, simply trading network width for depth alone does not automatically lead to better models. In the following sections, we highlight two deep learning architectures that we think deserve special attention and consideration for drug design and discovery: the Restricted Boltzmann Machine (RBM) and Convolutional Neural Networks (CNN). Many more deep learning approaches have been conceived, and the interested reader is referred to the respective literature.

3 Restricted Boltzmann Machine

Boltzmann machines are undirected, generative, stochastic neural networks that rose to prominence when Hinton and coworkers proposed contrastive divergence[60] as a fast unsupervised learning algorithm.[61] Restricted Boltzmann Machines (RBM) are instances of Boltzmann machines without intra- layer interactions (Figure 3). Boltzmann machines are

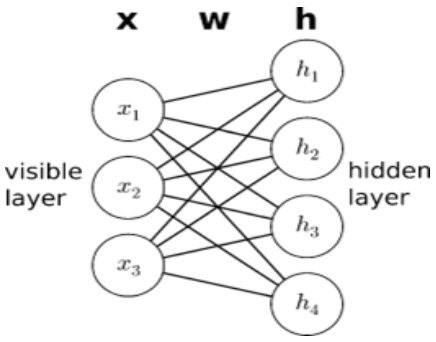

Figure 2. Schematic of a restricted Boltzmann machine. The visible layer is the interface between the network and the outside world.Input can be clamped to this layer, and the resulting visible neuron configuration represents a sample from the network's equi- librium distribution after an equilibration process. Hidden neurons(latent variables) transform the input data and detect prominent features. The hidden layer models the visible layer in conjunctionwith the weight matrix.

Not feed-forward models, which is a crucial difference from the deep feed-forward neural networks presented in Figure 2. The given input states serve as input to the hidden neurons but in turn are dependent on the resulting hidden neuron states that modify the "visible" neurons. This back and forth of mutual layer influencing eventually leads to an equilibrium of neuron activations regardless of the originally given input vector (Figure 3). The overall distribution governing the activation of visible neurons at one of these steps is the network's "belief" of how the visible neurons behave. The goal is to nudge the network's belief of the visible neuron state distribution into the direction of the actual input data distribution by intermittently adding new input data vectors and adjusting the weights in such a way that their probability is increased. The idea is that over time the believed visible state distribution eventually approximates the real state distribution, i.e. the network "learns" the input distribution in an unsupervised fashion. A popular criterion to evaluate the learning process is the reconstruction error, which is a measure of the ability of

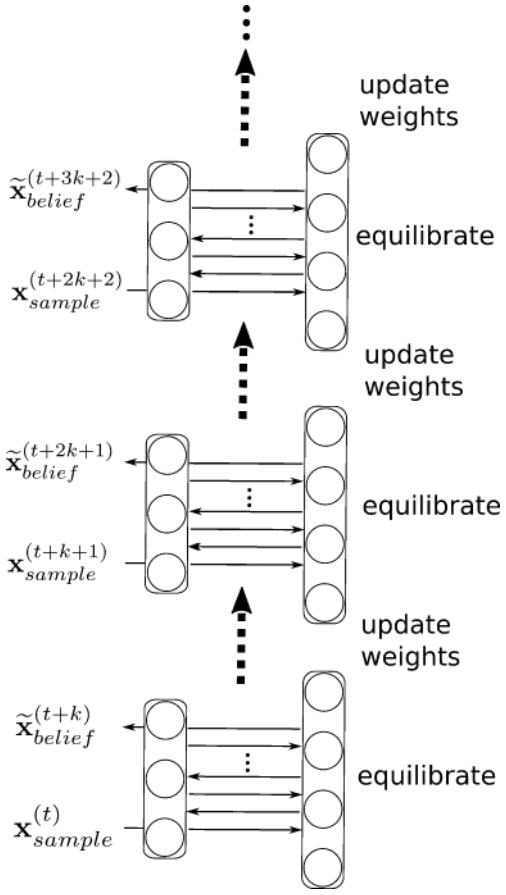

Figure 3. Restricted Boltzmann machine-learning. First, an input vector from the dataset enters the network. Subsequently, the hidden layer state is calculated with regard to the new input vector, from which anew input layer state can be calculated by the network. This iterative process continues for k steps, and the resulting visible layer state serves as a representative for the network's belief in the visible layer states. A weight update is calculated and the training cycle is re- iterated from the data input vector and the belief state vector.

4 Convolutional Neural Network

CNNs are inspired by neuroscience and imitate the image representation within the visual cortex that is created by successive transformations of the retinal signals in cortical areas called the "ventral stream" that end in the inferior temporal (IT) cortex.[68] Contemporary CNNs rival the IT-cortex's representational performance of a primate in "core visual object recognition".[69] The prototype of CNNs is the Neocognitron,[16] which was conceived in the early 1980s. Related architectures for phoneme recognition followed afew years later; thereafter, CNNs for object recognition in images and document reading were developed just prior to the turn of the millennium.[70] CNN improvement together with the general advances in deep learning became a focal point of interest when a deep CNN won the ILSVRC image recognition challenge in 2012

by a wide margin.[71] CNNs have dominated this challenge ever since, as well as the field of object recognition and detection in images in general. They now constitute the underlying architecture for a wide range of applications ranging from image understanding, video analysis, speech recognition and language understanding to artificial intelligence.[72] The two guiding principles of weight-sharing and local connectivity inaconvolutional net allow for the detection of features within topologically structured input data. Additional use of pooling and many layers enable the network to learn increasingly abstract features with each additional layer. The concept of weight sharing and local connectivity is depicted in Figure 5. Here, the input data corresponds to the input image, the filter corresponds tothe weight- connections, and the feature map (activation map) corresponds to the first hidden layer. In contrast to a fully connected network (Figure 2), hidden-layer neurons are only connect

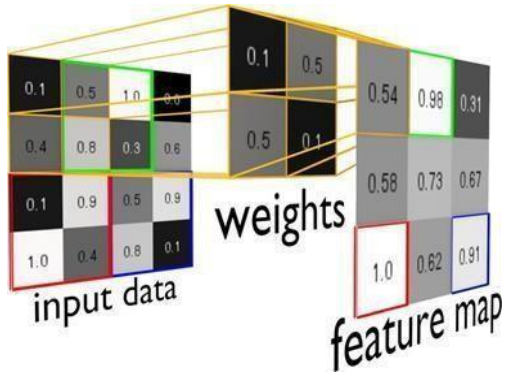

Figure 5. The figure depicts a toy example of a greyscale input image with pixel values between 0 and 1, a weight matrix (filter) and the resulting feature map. Pre-activations of hidden neuronswithin the feature map are obtained by weighting pixel-values within a 2 x2 receptive field with the corresponding weights of the filter and summing up the resulting four terms as shown in orange. To obtain the next feature map value, the filter is moved across the image with a stride of 1 (*e.g.,* from the orange to the green quadrant). The resulting feature map is a map of where thefeature encodedin thefilterhasoccurred in the input image. Here, the filter corresponds to a diagonal thatresults in astrong pre-acti-vation signal for the green, red and blue filter positions over the input image, which is reflected in the feature map accordingly.

6 Conclusion: Deep Learning to the Rescue?

Although the concepts of deep learning were introduced with the conception of multi-layer neural networks,[78] the life science community has adapted some of these techniques only recently (Table 1). In fact, neural networks have a long and fruitful history in drug discovery and design. Because they bear the risk of being easily over- trained and are perceived as a "black box", they have often been substituted by other approaches such as SVM models.[27,79] As noted by Winkler in a review article from 2004, continuous methodological advances in the field of neural networks alleviate some of the pitfalls and may have much to offer for hit and lead discovery.[80] Evidently, deep network architectures will require a particularly careful analysis and thorough definition of their respective domains of applicability,[81,82] and ideally provide hands-on guidelines for chemists. Appropriate confidence measures for neural network classifiers have been developed and can readily be ap plied.[83] Full appreciation of deep learning methods will likely only be achieved when the user gains an understanding of the underlying molecular features that enable pattern association and classification. With the development of new deep learning concepts such as RBMs and CNNs, the molecular modeler's tool box has been equipped with potentially game-changing methods. Judging from the success of recent pioneering studies, we are convinced that modern deep learning architecture will be useful for the coming age of big data analysis in pharmaceutical research, toxicity prediction, genome mining and chemogenomic applications, to name only some of the immediate areas of application. Personalized health care, in particular, could benefit from these capabilities. However, it may be wise not to put all of our eggs in one basket. We still need to fully understand the advantages and limitations of deep learning techniques. One should not blindly use them without simultaneously applying straightforward linear methods and pursuing chemical similarity approaches. By maintaining some healthy skepticism, we feel that the time is ripe for (re)discovering and exploring the usefulness of deep learning in drug discovery

Conflict of Interest

G. S. is a co-founder of inSili.com LLC, Zurich, and a consultant to the pharmaceutical industry.

Acknowledgements

This research was supported by the Swiss National Science Foundation (grant no. 200021_157190 and CR32I2_159737)

References

[1] a) R. Todeschini, V. Consonni, Molecular Descriptors for Cheminformatics, Wiley-VCH, Weinheim, 2009; b) R. Sawada, M. Kotera, Y. Yamanishi, Mol. Inf. 2014, 33, 719– 731.

[2] P. Willett, Mol. Inf. 2014, 33, 403 – 413.

[3] a) A. Lavecchia, Drug Discovery Today 2015, 20, 318–331; b) E. J. Gardiner, V. J. Gillet, J. Chem. Inf. Model. 2015, 55, 1781 – 1803.

[4] a) J. Devillers (ed.), Neural Networks in QSAR and Drug Design, Elsevier, Amsterdam; b) G. Schneider, P. Wrede, Prog. Biophys, Mol. Biol. 1998, 70, 175– 222; c) J. Zupan, J. Gasteiger, Neural Networks in Chemistry and Drug Design, Wiley-VCH, Weinheim, 1999; d) S. Agatonovic-Kustrin, R. Beresford, J. Pharm. Biomed. Anal. 2000, 22, 717–727.

[5] J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, V. Svetnik, J. Chem. Inf. Model. 2015, 55, 263– 274; b) J. Markoff, Scientists See Promise in Deep-Learning Programs. The New York Times, November 23, 2012.

[6] a) G. Schneider, S.-S. So, Adaptive Systems in Drug Design, Landes Bioscience, Georgetown, 2001;
b) M. Reutlinger, G. Schneider, J. Mol. Graph. Model. 2012, 34, 108– 117.

[7] G. E. Hinton, Cogn. Sci. 2014, 38, 1078– 1101.

[8] a) J. Gasteiger, T. Engel (eds.), Chemoinformatics: A Textbook, Wiley-VCH, 2003; b) G. Schneider, G. Downs (eds.), Machine Learning Methods in QSAR Modelling, QSAR Comb. Sci. 2003, 22(5); c) J. Bajorath (ed.) Chemoinformatics: Concepts, Methods, and Tools for Drug Discovery, Humana Press, Totowa, 2004; d) M. Dehmer, K. Varmuza, D. Bonchev (eds.) Statistical Modelling of Molecular Descriptors in QSAR/QSPR, Wiley-Blackwell, New York, 2011.

[9] a) G. Schneider, K.-H. Baringhaus, Molecular Design - Concepts and Applications. Wiley-VCH, Weinheim, 2008; b) G. Schneider (ed.), De Novo Molecular Design. Wiley-VCH, Weinheim, 2013.

[10] J. Gasteiger, Mini Rev. Med. Chem. 2003, 3, 789–796.

[11] Y. Bengio, Foundations and Trends in Machine Learning 2009, 2, 1– 127.

[12] a) G. E. Hinton, R. R. Salakhutdinov, Science 2006, 313, 504 – 507; b) Y. LeCun, Y. Bengio, G.Hinton, Nature 2015, 521, 436 – 444.

[13] Y. Roudi, G. Taylor, Curr. Opin. Neurobiol. 2015, 35, 110– 118.

[14] a) J. Fan, H. Liu, Adv. Drug Deliv. Rev. 2013, 65, 987–1000; b) D. C. Cires¸an, A. Giusti, L. M. Gambardella, J. Schmidhuber, Med. Image Comput. Comput. Assist. Interv. 2013, 16, 411 – 418; c) N. T. Issa, S. W. Byers, S. Dakshanamurthy, Expert Rev. Clin. Pharmacol. 2014, 7, 293– 298; d) F. F. Costa FF, Drug Discovery Today 2014, 19, 433– 440; e) S. J. Lusher, R. McGuire, R. C. van Schaik, C. D. Nicholson, J. de Vlieg, Drug Discovery Today 2014, 19, 859– 868; f) J. B. Brown, M. Nakatsui, Y. Okuno, Mol. Inf. 2014, 33, 732– 741; g) L. Richter, G. F. Ecker, Drug Discovery Today Technol. 2015, 14, 37 –41.

[15] Y. Park, M. Kellis, Nat. Biotechnol. 2015, 33, 825–826.

[16] a) K. Fukushima, S. Miyake, Pattern Recognit. 1982, 15, 455 – 469; b) K. Fukushima, Neural Netw. 2013, 37, 103– 119.

[17] a) P. J. Werbos, Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis, Harvard University, 1974; b) D. E. Rumelhart, G. E. Hinton, R. J. Williams, Nature 1986, 323, 533–536.

[18] N. Qian, T. J. Sejnowski, J. Mol. Biol. 1988, 202, 865–884.

[19] M. Punta, B. Rost, Methods Mol. Biol. 2008, 458, 203–230.

[20] M. Spencer, J. Eickholt, J. Cheng, IEEE/ACM Trans. Comput. Biol. Bioinform. 2015, 12, 103– 112.

[21] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R.E. Howard, W. Hubbard, LD. Jackel, Adv. Neural Inf. Proc. Sys. (NIPS) 1989, 2, 396–404.

[22] a) Y. Bengio, P. Simard, P. Frasconi, IEEE Trans. Neural Networks 1994, 5, 157– 166; b) S. Hochreiter, Y. Bengio, P. Frasconi, J. Schmidhuber, in: A Field Guide to Dynamical Recurrent Neural Networks (S. C. Kremer, J. F. Kolen, eds.), IEEE Press, New York, 2001.

[23] a) G. E. Hinton, Trends Cogn. Sci. 2007, 11, 428– 434; b) G. E. Hinton, S. Osindero, Y. W. The, Neural Comput. 2006, 7, 1527– 1554.

[24] a) G. Schneider, J. Schuchhardt, P. Wrede, Comput. Appl. Biosci. 1994, 10, 635– 645; b) G. Schneider, Neural Netw. 2000, 13, 15 –16; c) S. Agatonovic-Kustrin, R. Beresford, J. Pharm. Biomed. Anal. 2000, 22, 717– 727.

[25] a) L. Breiman, Machine Learning 2001, 45, 5 – 32; b) N. Meinshausen, J. Mach. Learn. Res. 2006, 7, 983–999.

[26] a) Y. Sakiyama, Expert Opin. Drug Metab. Toxicol. 2009, 5, 149– 169; b) D. Plewczynski, S. A. Spieser, U. Koch, Comb. Chem. High Throughput Screen. 2009, 12, 358– 368; c) B. Sprague, Q. Shi, M. T. Kim, L. Zhang, A. Sedykh, E. Ichiishi, H. Tokuda, K. H. Lee, H. Zh, J. Comput. Aided Mol. Des. 2014, 28, 631– 646; d) M. A. Khamis, W. Gomaa, W. F. Ahmed, Artif. Intell. Med. 2015, 63, 135–152; e) T. Rodrigues, D. Reker, M. Welin, M. Caldera, C. Brunner, G. Gabernet, P. Schneider, B. Walse, G. Schneider, Angew. Chem. Int. Ed. 2015, in press.

[27] a) E. Byvatov, U. Fechner, J. Sadowski, G. Schneider, J. Chem. Inf. Comput. Sci. 2003, 43, 1882–1889; b) E. Byvatov, G. Schneider, Appl. Bioinformatics 2003, 2, 67 –77.

[28] a) L. Yang, P. Wang, Y. Jiang, J. Chen, J. Chem. Inf. Model. 2005, 45, 1804– 1811; b) I. I. Baskin, V. A. Palyulin, N. S. Zefirov, Methods Mol. Biol. 2008, 458, 137–158.

[29] A. Givehchi, G. Schneider, Mol. Divers. 2005, 9, 371–383.

[30] a) A. C. Tsoi, A. D. Back, IEEE Trans. Neural Netw. 1994, 5, 229 – 239; b) B. Hammer, A. Micheli, A. Sperduti, M. Strickert, Neural Netw. 2004, 17, 1061–1085; c) D. V. Buonomano, Neuron 2009, 63, 423– 425; d) D. Plewczynski, J. Mol. Model. 2011, 17, 2133– 2141.

BIOGRAPHIES

Kallesh SC

Studying engineering in BIET, Davanagere -577004

Khushi DN

Siddesh DS

Thejaswi MR