J Biochem Technol (2023) 14(1): 1-6
https://doi.org/10.51847/iW1DfVoXVw
ISSN: 0974-2328

Applications of Deep Learning and Machine Learning in Computational Medicine

Rama Adiga*, Titas Biswas, Perugu Shyam

Received: 06 December 2022 / Received in revised form: 20 February 2023, Accepted: 23 February 2023, Published online: 15 March 2023

Abstract

Computational medicine has emerged due to the advances in medical technology in parallel with big data and artificial intelligence. A new way of treating complex diseases is evolving called 'Precision Medicine' fueled by big data extracting meaningful information from individual variability. At the forefront is biomedical research aiming to promote the area of precision medicine. Though traditional machine learning methods have built successful models for cancer diagnosis to sars-cov2 pulmonary infection, the advent of modern deep learning methods has had phenomenal growth in genomics, electronic health records, and drug development. The challenges in Deep learning applications in medicine include lack of data, privacy, heterogeneity of data, and interpretability. Analysis and discussion on these problems provide a reference to improve the application of deep learning in medical health.

Keywords: Computational medicine, Deep learning, Genomics, Big data

Introduction

Computational Medicine comprises many interdisciplinary subjects such as medicine, biology, mathematics, and computer science. Artificial intelligence methods are applied to understand disease mechanisms in humans by utilizing big data analytics to help in disease prediction and guiding clinical decision-making.

The pharmaceutical industry has particularly benefitted from its utility in drug discovery and clinical research including strategizing in the early and successful development of new drugs. Computational medicine can cut down the time of drug development to an average of one to two years. The field of intelligent computing is transforming health care delivery and medical practice.

Rama Adiga*, Titas Biswas
Nitte (Deemed to be University), Nitte University Centre for Science Education & Research (NUCSER), Division of Bioinformatics and Computational Genomics, Deralakatte, Paneer Campus, Mangalore, India 575018.

Perugu Shyam
Deptment of Biotechnology, Chemical, and Biotechnology Block, National Institute of Technology, Warangal (NITW), 506004, Telangana, India.

***E-mail:** rama_adiga@nitte.edu.in

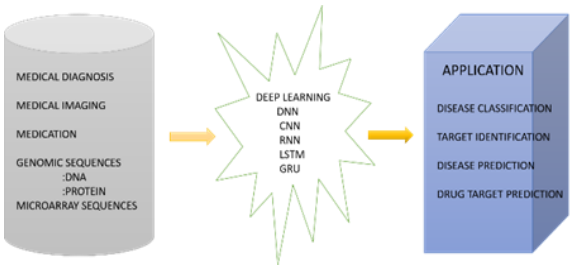


Figure 1. Deep learning Applications in Computational Medicine

The key challenges in effective data mining of biomedical information are that they are often multi-dimensional which implies that pre-processing involves primarily reducing the dimension. The datasets are also small and interspersed with noise, making them un-interpretable and difficult to process. It is desirable to consider the criterion while developing protocols and methods for handling biomedical data. Deep learning as a branch of machine learning has been fraught with superior abilities in training large datasets for artificial intelligence (Han *et al*., 2018). Several authors have developed effective methods of deep learning in areas like computer vision (He *et al.*, 2016), Speech recognition (Abdel-Hamid *et al*., 2014), and natural language processing (NLP). In the traditional machine learning domain and experts extract features from datasets referred to as 'feature engineering' and utilize the features to construct machine learning models for analysis. Common methods like random forest and support vector machine are applied for the analysis. The manual extraction of data may lead to bias which is inefficient at performing tasks and failure in achieving a high-performance model.

Deep learning or deep neural networks (Goodfellow *et al*., 2014; LeCun *et al*., 2015; Goodfellow *et al*., 2016) has several success stories of artificial intelligence. Inspired by neuroscience, the architectural principle of Deep Neural Networks (DNN) and Analog Neural networks (ANN) have been used which primarily involve non-linear layers of nodes similar to the working in the brain (Fukushima et al., 1988; Riesenhuber and Poggio, 1999; Schmidhuber *et al*., 2015; He *et al*., 2016; Silver *et al*., 2016). Neural networks comprise several layers of neurons with connections between the layers to be able to convert the input to the output. The high-level network enables faster and more advanced learning which traditional machine learning is unable to handle efficiently. A recent trend is to apply deep learning to biomedical data analysis as it is one of the advanced deep learning methods. **Figure 1** represents the processes and outcomes of using biomedical data. However, the challenges faced by data analysts

© 2023 The Author(s). This is an **Open Access** article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0).
https://creativecommons.org/licenses/by/4.0/deed.en

are that the data is not easily accessible for processing and not easily processible since data is unclean. If these constraints were met what are the applications of deep learning in the upcoming field of Genomics, Drug-development and Electronic Health Records?

Deep Learning Applications in Genomics

The study of structural and functional aspects of the gene is Genomics. Other applications of genomics are genomic editing which involves targeting and modification of genomic sequences. Genome editing for genetic deficiency disorders to improve public health has been carried out under stringent guidelines set by various scientific eg. WHO (WHO guidelines, 2022). Gene expression studies are also popular to identify various organisms which live symbiotically in the human body as well as disease-causing organisms. The field of genomics has under its code hidden patterns which require advanced deep learning methods to extract meaningful information. Biological databases are at the forefront providing vital information free to the public eg. NCBI, EMBL, etc. The following paragraph explains a meta-analysis presented using publicly available data in NCBI (GEO dataset).

Meta-Analysis of the SARS-CoV2 Dataset

The unsupervised learning method (kNN and SOM) was applied to the dataset from GEO (NCBI) for classifying the dataset without labeling. The pulmonary tissue infected with the SARS-CoV2 dataset was available. Remarkable heterogeneity in expression levels of RNA as well as the spatial location has been reported (Desai *et al*., 2020). Neural network models have become powerful areas for development in the area of machine learning. Using SOM maps are popular algorithms for reducing dimensions with neural networks which can directly be compared.

Method: Gene expression data were downloaded from the GEO database of NCBI, using the following search: ("SARS CoV-2" AND "expression profiling" AND "spatial heterogeneity"). The following datasets were obtained having accession GSE159787 and GSE159785. 726 sample datasets were available in a set of two experiments with autopsy tissue from patients who had succumbed to SARS-CoV2. Gene expression of immune response genes was studied across the spectrum of high and low viral load regions of the tissue. About 48 segments were available within the same patient from a total of six infected patients which was useful in studying the heterogeneity within the tissue (Desai *et al*., 2020). The dataset contained expression data from six infected patients which were collected from cadaver patients and stored freeze-dried.

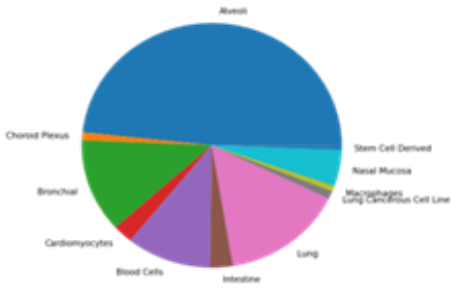


Figure 2. Proportion of type of infected tissue used in the analysis

The datasets **(Figure 2)** were organized into tissue sub-structure namely alveoli, bronchial, and Ln type as well as segment-wise to develop a meaningful relationship with each other. After some data exploration, the samples were largely classified into geometric, PanCK-Neg, or PanCK-Pos by the segment type as given in the original nomenclature into a Venn diagram **(Figure 3)**. Also, the cells were alveolar, bronchial or LN type by the tissue substructure. Thus, we could identify the shared properties among the substructures by the segment type of individual samples. The weighted Venn and unweighted Venn are depicted in **Figure 3**. For better visualization, we used heat maps from the seaborn library of machine learning methods **(Figure 4)**.

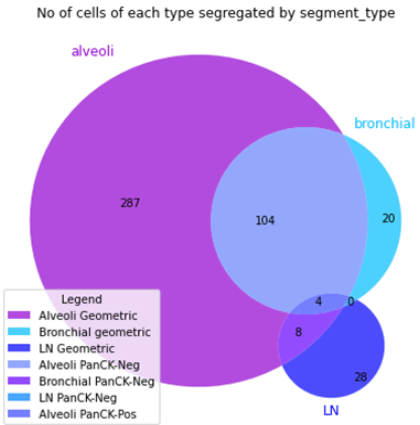


Figure 3. Number of cells of each type segregated by segment type (weighted Venn)

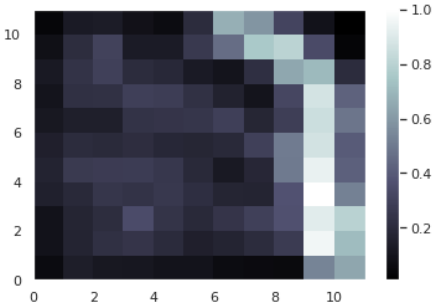


Figure 4. In the exploratory data analysis, classification of the fields in the dataset, without explicitly labeling them

Powerful machine learning methods, hold great promise to provide biological insights from large and often heterogeneous data. Two popular deep learning methods, convolutional neural network (CNN) and recurrent neural network (RNN), are used in genomics. Every complex question in biology will need specific machine learning approaches, e.g., support vector machine (SVM) and random forest, and several other combinations.

Deep learning in genomics helps in reducing high-dimensional features into easily interpretable information and unraveling hidden features in human disease. Other genomic analyses where deep learning has benefitted example binding site prediction of RNA binding proteins (Zeng *et al*., 2016), gene expression abundance studies (Washburn *et al*., 2019; Agarwal & Shendure, 2020). Use of CNN in predicting functional activity from genomic data. To understand yeast gene expression using microarrays, Chen *et al*. (2016) used autoencoders (Chen *et al*., 2016; Govindaraj *et al*., 2022). Others have used Deep CNN model with success in understanding genomics (Zhou *et al*., 2018; Yuan & Joseph, 2019; Gao *et al*., 2020). A combination of one-dimensional CNN, RNN has been very popular in extracting meaningful information. While the former extracts feature reducing the dimension the latter has been used to discover patterns from genomic sequence data. NLP has also been applied to gather data in innovative ways. Multimodal learning is currently popular using data from varied sources offering unique possibilities in genomic data mining. To take an example of combining gene sequence data with electronic health records and imaging records.

Major pitfalls in genomic data mining are insufficient data for efficient learning. If data is insufficient, it is often difficult to find a suitable deep-learning model for the task. This may lead to poor outcomes. For effective training to take place the dataset must be large.

Guidelines to solve the problem of lack of data in genomics: 1. To use data enhancement to expand the data layer and sample size.2. Use dropout methods to improve model ability and enhance model performance. 3. To use transfer technology to train the model on a larger dataset that is unrelated and to reuse it on the desired set with adjusted parameters.

Electronic Health Records (EHR)

Similar to genomics prediction, Computational Medicine uses the following neural networks for processing one-dimensional CNN, RNN, GRU, and LSTM along with NLP.

Storage of electronic health records is generally in a format that is structured as well as unstructured. The former includes treatment information on patient demographic and diagnosis information as well as laboratory test results (Jensen *et al*., 2012).

One of the advantages of mining electronic health records is the delivery of timely treatment by early prediction of disease onset. It can also analyze disease-to-drug relationships which can be vital to the clinician in decision-making Dirican (2023).

The records are generally subject-coded using medical terminology called ontologies. Such terms are extensively used to describe conditions, drugs, or even diagnostic processes. An umbrella code for the entire medical concept is a highly specialized field beyond the realm of data scientists to simplify the relationship required to code the condition of the patient. Another difficulty is extracting meaningful relationships that reflect the concept in the code. Research in the area is geared towards mapping clinical terminology to the relevant code to convert the high dimensionality into low dimension and transform them for embedding. A complex relationship is encountered when determining patient mortality and readmission which is an indicator of quality care in hospitals. Deep learning methods can represent correctly the non-linear relationship through hidden layers of neural networks (Miotto *et al*., 2016; Mohandas *et al*., 2022).

Drug Development

With drug development reaching new heights, deep learning has immense application in the field of drug research. With typical drug development taking approx. 10 years is a very resourceful and time-consuming process. Deep learning can help in studying the interaction between a drug and its target. This step can help cut down the time signature of bringing the drug to the market. Prediction of binding affinity between drug and target (Wen *et al*., 2017; Wang *et al*., 2018; Hu *et al*., 2019) has been possible with the use of autoencoders to study drug-target interaction. Others used LSTM on the feature space to predict the interaction between the drug and the target. Zhang *et al*. (2020) developed a deep-learning model for target recognition and drug reuse by learning low-dimensional vector representations of drugs and targets.

A unique method of combining sequence data with ligand fingerprints was proposed using CNN and graph neural network (Tsubaki, *et al*., 2019). It excelled when compared to techniques like kNN, random forest, and SVM. The databases for extracting information from drug compounds are DrugBank, KEGG, and PubChem. LSTM is a special recurrent neural network (RNN) giving superior results as compared to previous methods since it uses memory blocks for storing temporal states.

Table 1. Models developed and the author's contribution is listed.

S/N	Model name	Description	Reference
1.	Word2vec model	Skip-gram method representation of disease and clinical concepts	(Mikolov *et al*., 2013)
2.	RNN	Used for predicting hospitalization	(Zhang J. *et al*., 2018)
3.	Three-layer stacked denoising autoencoder	Unsupervised deep learning method to capture the hierarchical relationship	(Miotto *et al*., 2016)
4.	RNN & CNN	To extract patterns in patient information	(Ma *et al*., 2018)

5.	Long Short-Term Memory (LSTM)	Learning based on a single-layer decision tree	(Rajkomar *et al*., 2018)
6.	Time-aware neural network models	The deep dynamic memory model	(Pham *et al*., 2016)
7.	Gated Recurrent Unit (GRU)	For accurate segmentation of 3D data by applying masking & time-interval	(Che *et al*., 2018)

Challenges

Some of the challenges to Deep learning in Computational medicine and certain solutions are discussed here. Learning is not simple and sufficient data is essential for efficient Deep learning. In a healthcare setting medical datasets need to be handled differently as the patient number and the disease category are unevenly matched.

To solve the model's poor generalization ability, the dropout method is sometimes used. Other methods include data enhancement which is achieved by translation, clipping and scaling for new image generation for creating a robust model.

Transfer learning is also a useful method where data learning happens in another related task with sufficient data with model parameters being fine-tuned before finally utilizing it (Shin *et al*., 2016). Data integration is used by some workers to improve the model's ability (Dai *et al*., 2018).

Another logical solution is combining domain expertise knowledge with image information for efficient training of deep learning.

Model Interpretability is an important issue to translate the work to the clinician and be able to make a sound clinical decision based on the prediction and patient satisfaction.

The interpretability of models is very important. Because if a model can provide sufficient and reliable information, physicians will trust the results of the model leading to correct and appropriate decisions; at the same time, an interpretable model can also provide a comprehensive understanding of the progress of patients.

Table 2. Drug development Models developed using Deep learning and their contributions are listed.

	Model name	Description	Availability
1	Torch Drug	Torch Drug is a PyTorch-based machine learning toolbox designed for several purposes.	https://github.com/DeepGraphLearning/torchdrug
2	Drug Explorer (Drug Ex v3)	Scaffold-Constrained Drug Design with Graph Transformer-based Reinforcement Learning	https://github.com/XuhanLiu/DrugEx
.	Deep Mol	Smoother approach to many drug discovery and chemoinformatics problems. It uses Tensor flow, Keras, Scikit-learn, and DeepChem	https://github.com/BioSystemsUM/DeepMol
4	Deep Chem	Deep learning in drug discovery, materials science, quantum chemistry, and biology.	https://github.com/deepchem/deepchem
5	Deep Conv-DTI	Combination of protein sequences and molecular fingerprints of ligands to generate a fully connected layer	Lee *et al*. (2019)
.	Deep DTA	Deep drug-target binding affinity prediction	https://github.com/hkmztrk/DeepDTA

Conclusion

Discussion of various deep learning methods which may be used in computational medicine, among them the prominent field are genomics, electronic health record, and drug development.

Biomedical data mining can be of immense use in clinical decision-making if the methods are reliable, interpretable, and standardized. Ideas for improvement were provided for computational researchers in the development of models in the health field.

Acknowledgments: The author wishes to thank Dr. Anirban Chakraborty, Director of Nitte University Centre for Science Education and Research (NUCSER), and the management of Nitte (Deemed to be University), Deralakatte, Mangalore, Karnataka, India for supporting in research including the present work.

Conflict of interest: None

Financial support: None

Ethics statement: None

References

Abdel-Hamid, O., Mohamed, A. R., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *22*(10), 1533-1545. doi:10.1109/TASLP.2014.2339736

Agarwal, V., & Shendure, J. (2020). Predicting mRNA abundance directly from genomic sequence using deep convolutional neural networks. *Cell Reports*, *31*(7), 107663. doi:10.1016/j.celrep.2020.107663

Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, *8*(1), 6085. doi:10.1038/s41598-018-24271-9

Chen, Y., Li, Y., Narayan, R., Subramanian, A., & Xie, X. (2016). Gene expression inference with deep learning. *Bioinformatics*, *32*(12), 1832-1839. doi:10.1093/bioinformatics/btw074

Dai, L., Fang, R., Li, H., Hou, X., Sheng, B., Wu, Q., & Jia, W. (2018). Clinical report guided retinal microaneurysm detection with multi-sieving deep learning. *IEEE Transactions on Medical Imaging*, *37*(5), 1149-1161. doi:10.1109/tmi.2018.2794988

Desai, N., Neyaz, A., Szabolcs, A., Shih, A. R., Chen, J. H., Thapar, V., Nieman, L. T., Solovyov, A., Mehta, A., Lieb, D. J., et al. (2020). Temporal and spatial heterogeneity of host response to SARS-CoV-2 pulmonary infection. Nature *Communications*, *11*(1), 6319. doi:10.1038/s41467-020-20139-7

Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Networks*, *1*(2), 119-130.

Gao, D., Morini, E., Salani, M., Krauson, A. J., Ragavendran, A., Erdin, S., Logan, E. M., Chekuri, A., Li, W., Dakka, A., et al. (2020). A deep learning approach to identify new gene targets of a novel therapeutic for human splicing disorders. *BioRxiv*, 2020-02.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (vol. 1) cambridge.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley D., & Ozair, S. (2014). Generative adversarial nets," in Advances in Neural Information Processing Systems 27, eds Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Montreal, QC: NIPS), pp. 2672-2680.

Govindaraj, A., Paulpandian, S. S., & Shanmugam, R. (2022). A comparative assessment of the impact of *Citrullus lanatus* rind and pulp extracts on *streptococcus mutans*. *International Journal of Dental Research and Allied Sciences*, *2*(2), 1-6. doi:10.51847/ffD4G8LgXK

Han, S. S., Park, G. H., Lim, W., Kim, M. S., Na, J. I., Park, I., & Chang, S. E. (2018). Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network. *PloS one*, *13*(1), e0191493. doi:10.1371/journal.pone.0191493

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Hu, P., Huang, Y. A., You, Z., Li, S., Chan, K. C., Leung, H., & Hu, L. (2019). Learning from deep representations of multiple networks for predicting drug–target interactions. In *Intelligent Computing Theories and Application: 15th International Conference, ICIC 2019, Nanchang, China, August 3–6, 2019, Proceedings, Part II 15* (pp. 151-161). Springer International Publishing. doi:10.1007/978-3-030-26969-2_14

Jensen, P. B., Jensen, L. J., & Brunak, S. (2012). Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*, *13*(6), 395-405. doi:10.1038/nrg3208

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. doi:10.1038/nature14539

Ma, T., Xiao, C., & Wang, F. (2018, May). Health-atm: A deep architecture for multifaceted patient health record representation and risk prediction. In *Proceedings of the 2018 SIAM International Conference on Data Mining* (pp. 261-269). Society for Industrial and Applied Mathematics.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, *6*(1), 1 26094.

Mohandas, R., Ramani, P., & Mohapatra, S. (2022). Exploring Corono-condylar distance as a radiographic marker for chronological age. *International Journal of Dental Research and Allied Sciences*, *2*(2), 7-9. doi:10.51847/xF069fnRvk

Pham, T., Tran, T., Phung, D., & Venkatesh, S. (2016). DeepCare: A deep dynamic memory model for predictive medicine. In Advances in Knowledge Discovery and Data Mining. PAKDD 2016. Lecture Notes in Computer Science, eds J. Bailey, L. Khan, T. Washio, G. Dobbie, J. Huang, and R. Wang (Cham: Springer).

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj N., & Hardt, N. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, *1*(18).

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019-1025.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85-117. doi:10.1016/j.neunet.2014.09.003

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, *35*(5), 1285-1298. doi:10.1109/tmi.2016.2528162

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484-489. doi:10.1038/nature16961

Tsubaki, M., Tomii, K., & Sese, J. (2019). Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics*, *35*(2), 309-318. doi:10.1093/bioinformatics/bty535

Wang, L., You, Z. H., Chen, X., Xia, S. X., Liu, F., Yan, X., Zhou, Y., & Song, K. J. (2018). A computational-based method for predicting drug–target interactions by using stacked autoencoder deep neural network. *Journal of Computational*

Biology, *25*(3), 361-373. doi:10.1089/cmb.2017.0135.

Washburn, J. D., Mejia-Guerra, M. K., Ramstein, G., Kremling, K. A., Valluru, R., Buckler, E. S., & Wang, H. (2019). Evolutionarily informed deep learning methods for predicting relative transcript abundance from DNA sequence. *Proceedings of the National Academy of Sciences*, *116*(12), 5542-5549. doi:10.1073/pnas.1814551116

Wen, M., Zhang, Z., Niu, S., Sha, H., Yang, R., Yun, Y., & Lu, H. (2017). Deep-learning-based drug–target interaction prediction. *Journal of proteome research*, *16*(4), 1401-1409.

WHO guideline (accessed May 5th, 2022) Available from: https://www.who.int/news/item/12-07-2021-who-issues-new-recommendations-on-human-genome-editing-for-the-advancement-of-public-health# N.

Yuan, Y., & Bar-Joseph, Z. (2019). Deep learning for inferring gene relationships from single-cell expression data. *Proceedings of the National Academy of Sciences*, *116*(52), 27151-27158. doi:10.1073/pnas.1911536116

Zeng, H., Edwards, M. D., Liu, G., & Gifford, D. K. (2016). Convolutional neural network architectures for predicting DNA–protein binding. *Bioinformatics*, *32*(12), i121-i127. doi:10.1093/bioinformatics/btw255

Zhang, H., Saravanan, K. M., Yang, Y., Hossain, M. T., Li, J., Ren, X., Pan, Y., & Wei, Y. (2020). Deep learning based drug screening for novel coronavirus 2019-nCov. *Interdisciplinary Sciences: Computational Life Sciences*, *12*, 368-376. doi:10.1007/s12539-020-00376-6 2020

Zhang, J., Kowsari, K., Harrison, J. H., Lobo, J. M., & Barnes, L. E. (2018). Patient2vec: A personalized interpretable deep representation of the longitudinal electronic health record. *IEEE Access*, *6*, 65333-65346. doi:10.1109/access.2018.2875677

Zhou, J., Theesfeld, C. L., Yao, K., Chen, K. M., Wong, A. K., & Troyanskaya, O. G. (2018). Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nature Genetics*, *50*(8), 1171-1179. doi:10.1038/s41588-018-0160-6\