

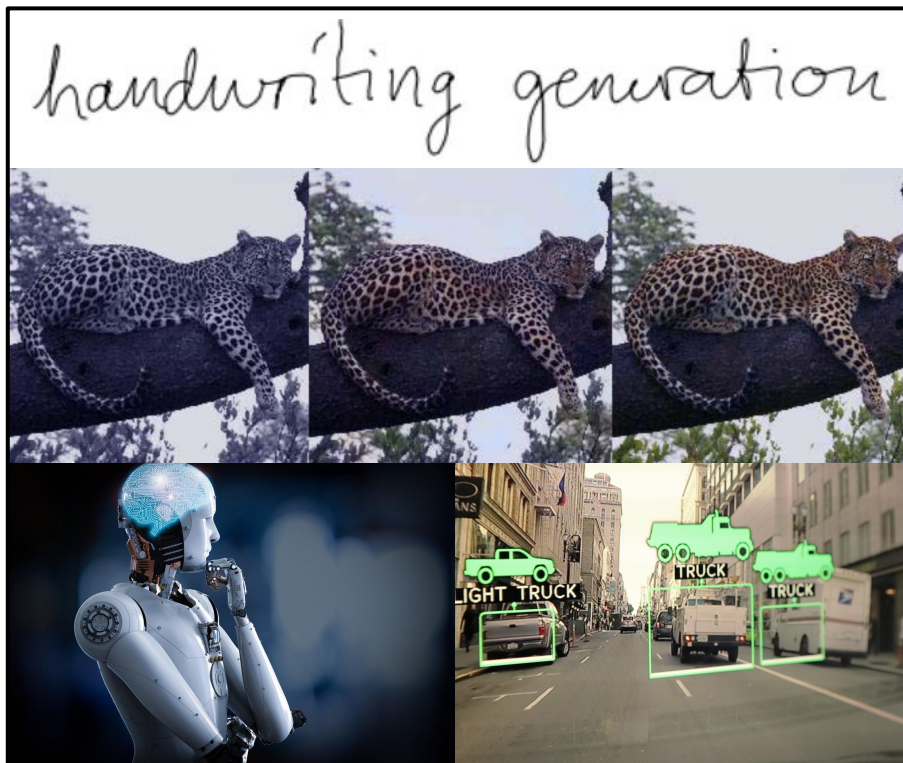
# Deep Learning

*in Biomedical Data Science*



# Applications

- Image Color Restoration
- Image Captioning
- Object Classification
- Robotics
- Self-driving Cars
- Machine Translation
- Speech Recognition
- Text Generation
- Handwriting Generation
- Music Composition
- & more



# Applications



OpenAI

@OpenAI

Follow

We've trained an unsupervised language model that can generate coherent paragraphs and perform rudimentary reading comprehension, machine translation, question answering, and summarization — all without task-specific training:

[blog.openai.com/better-language...](https://blog.openai.com/better-language-model/)



We've trained a large-scale unsupervised language model which generates coherent paragraphs of text, achieves state of the art performance on many language modeling benchmarks, and performs rudimentary reading comprehension, machine translation, question answering, and summarization —



0:38 312K Views

9:03 AM - 14 Feb 2019

2,782 Retweets 6,603 Likes



DeepMind's new AI just beat top human pro-gamers at Starcraft II for the first time

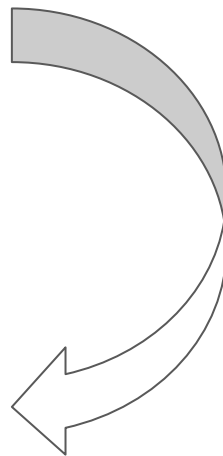


[\[https://thispersondoesnotexist.com\]](https://thispersondoesnotexist.com)

# One-Hot Encoding of Sequences

C C A A T T G T T T T G

	C	C	A	A	T	T	G	T	T	T	T	G
A	0	0	1	1	0	0	0	0	0	0	0	0
C	1	1	0	0	0	0	0	0	0	0	0	0
T	0	0	0	0	1	1	0	1	1	1	1	0
G	0	0	0	0	0	0	1	0	0	0	0	1



# Applications

## Biomedical Data Science

Fast and Deep  
with predic

Sepp Hochreiter, Pietro Di Lorenzo

Bioinformatics Bioinformatics

<https://doi.org/10.1093/bioinformatics/btw255>

Published

Split View

Abstract

Motivation: Classification of protein sequences

2007

Deep prediction

Pietro Di Lorenzo

Bioinformatics

<https://doi.org/10.1093/bioinformatics/btw255>

Published

Split View

Abstract

Motivation: Contact prediction

2012

RESEARCH ARTICLE  
The human genetic

Hui Y. Xiong<sup>1,2</sup>  
+ See all authors

Science 09 Jan 2015  
Vol. 347, Issue 6200  
DOI: 10.1126/science.1262117

Article

You are currently viewing

Prediction

Most eukaryotic genomes contain large regions of uninterruptedly regulated tissues. F. Xiong et al.

2015

Methodology  
Deep gene

Tao Zeng, R

BMC Bioinformatics

<https://doi.org/10.1186/s12859-015-0688-1>

Received: 2

Abstract

Background

Profiling scales of *in situ* of the

2015

Predicting DNA

Babak Ali

Nature Biotechnology

Abstract

Known biological processes show

2015

## Convolutional neural network architectures for predicting DNA–protein binding

Haoyang Zeng, Matthew D. Edwards, Ge Liu, David K. Gifford

Bioinformatics, Volume 32, Issue 12, 15 June 2016, Pages i121–i127,

<https://doi.org/10.1093/bioinformatics/btw255>

Published: 11 June 2016

Split View

Cite

Permissions

Share

### Abstract

**Motivation:** Convolutional neural networks (CNN) have outperformed conventional methods in modeling the sequence specificity of DNA–protein binding. Yet inappropriate CNN architectures can yield poorer performance than simpler models. Thus an in–depth understanding of how to match CNN architecture to a given task is needed to fully harness the power of CNNs for computational biology applications.

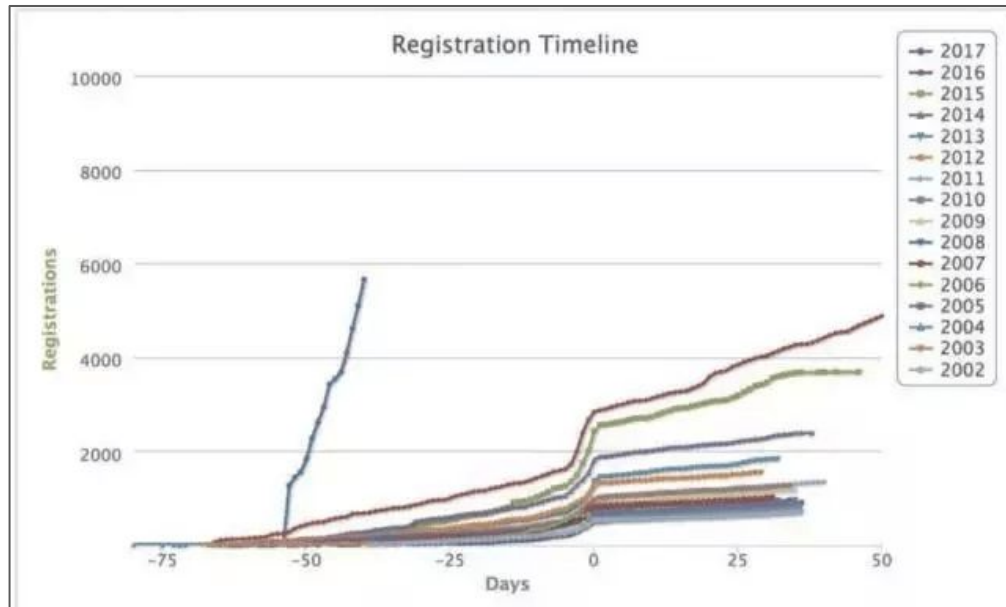
**Results:** We present a systematic exploration of CNN architectures for predicting DNA sequence binding using a large compendium of transcription

2016

# Applications

## *Biomedical Data Science*

- DeepBind, DeepFold, DeepSEA, DeepCNF, DeepVariant, DeepMotif, DeepLNC, DeepCpG, DeepCyTOF, DeepChrome, DeepWAS, DeepSplice, and other methods

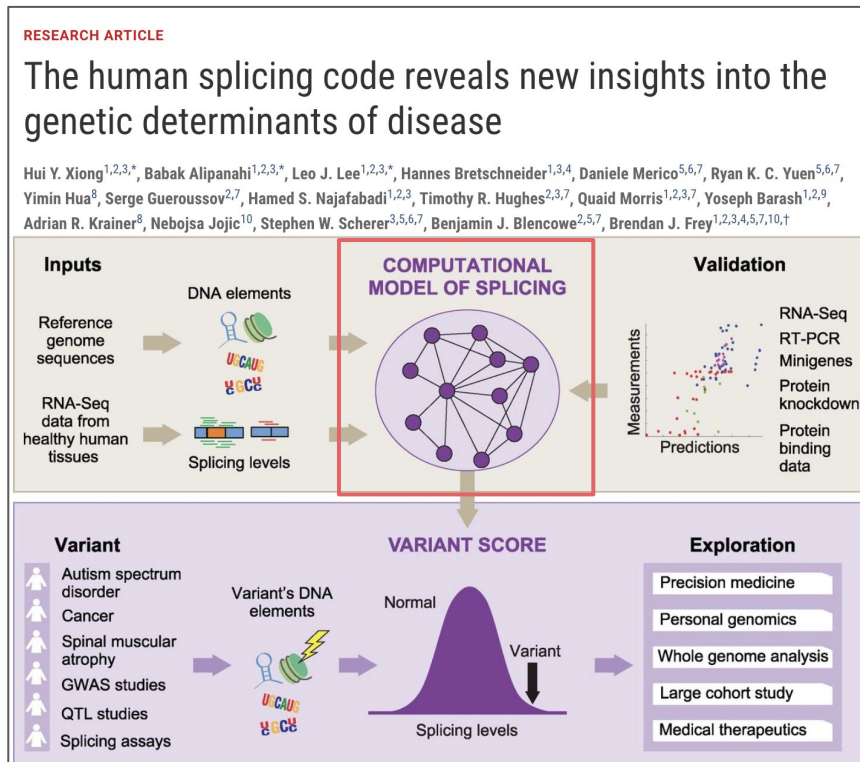


NeurIPS Registration Timeline [2002-17] @lxbrun

# Applications

## Biomedical Data Science

- Prediction of RNA Splicing
- 1393 sequence features as input
- Fully-connected neural networks
- RNA-seq from *ILLUMINA Human Body Map 2.0* project
- Assessment of >650,000 SNPs effect on splicing regulation from the NCBI database of genetic variation (*dbSNP*) and Human Gene Mutation Database (*HGMD*)



# Applications

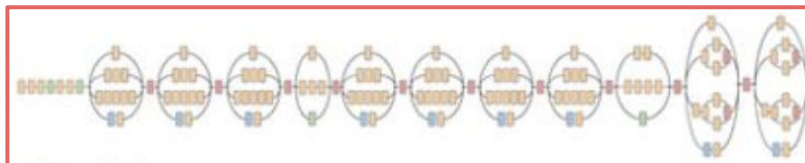
## Biomedical Data Science

- Skin cancer classification
- 129,450 clinical images from public

## Dermatologist-level classification of skin

Skin lesion image

Deep convolutional neural network (Inception v3)



- Convolution
- AvgPool
- MaxPool
- Concat
- Dropout
- Fully connected
- Softmax

Training classes (757)

- Acral-lentiginous melanoma
- Amelanotic melanoma
- Lentigo melanoma
- ...
- Blue nevus
- Halo nevus
- Mongolian spot
- ...
- 
- 
- 

Inference classes (varies by task)

- ⊕ ● 92% malignant melanocytic lesion
- ⊕ ○ 8% benign melanocytic lesion

[2017]



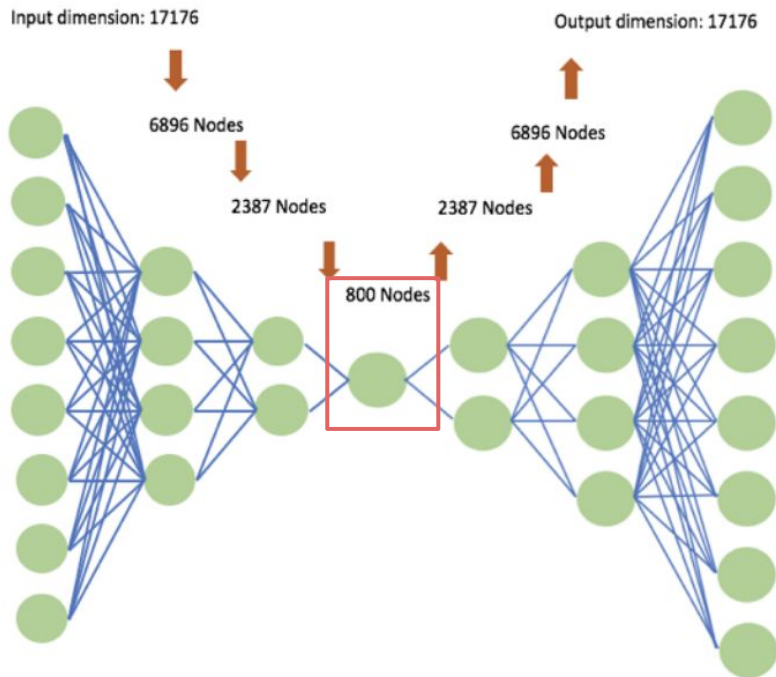
# Applications

## Biomedical Data Science

- Denoising Autoencoder (AE) for genomic data imputation
- $|I| = |O|$ , bottleneck layer  $|B| \ll |I|, |O|$
- 17,000+ genes, 269,000+ CpG sites
- Pan-Cancer RNA-seq and DNA Methylation data from The Cancer Genome Atlas (**TCGA**)

A deep learning framework for imputing missing values in  
genomic data

Yeping Lina Qiu<sup>1,2</sup>, Hong Zheng<sup>2</sup>, Olivier Gevaert<sup>2,3,\*</sup>



[2018]

# Challenges

- **Interpretability**
  - Different from interpretability in Optimization
  - Weight matrix interpretation
  - Input feature prioritization
  - Biologically or clinically-driven
  - Earlier and later layers usually have more importance
- **Data availability**
- **Hyperparameter Optimization**
  - Set *before* training
  - Finding suboptimal combinations can be time consuming



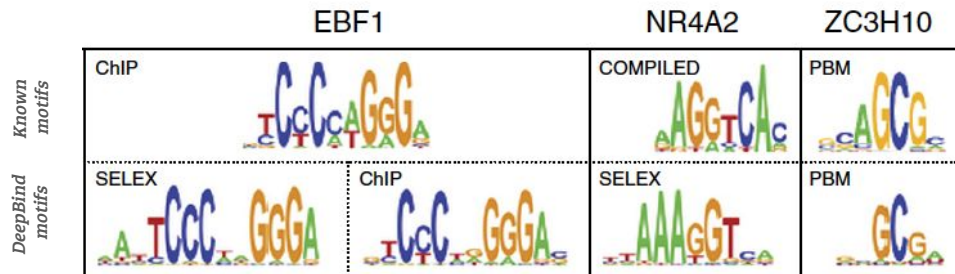
# Interpretability

## DeepBind

- Interpretability
  - Different from interpretability in Optimization
  - **Weight matrix interpretation**
  - Input feature prioritization
  - Biologically or clinically-driven
  - Earlier and later layers usually have more importance
- What would happen if the number of convolution filters is quite high?

Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning

Babak Alipanahi<sup>1,2,6</sup>, Andrew Delong<sup>1,6</sup>, Matthew T Weirauch<sup>3-5</sup> & Brendan J Frey<sup>1-3</sup>



# Interpretability

## SurvivalNet

- Interpretability
  - Different from interpretability in Optimization
  - Weight matrix interpretation
  - **Input feature prioritization**
  - Biologically or clinically-driven
  - Earlier and later layers usually have more importance
- 17,000-18,000 features dominated by gene expression and clinical variables
- Why would features be prioritized? What would potential downstream analyses be?

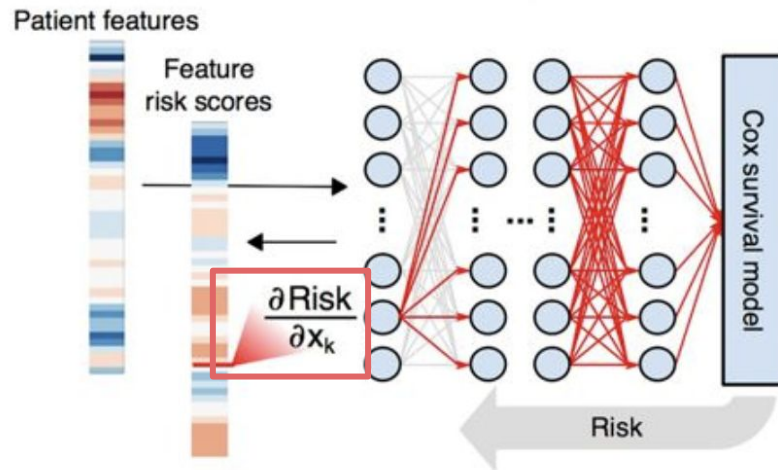
## SCIENTIFIC REPORTS

OPEN

Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models

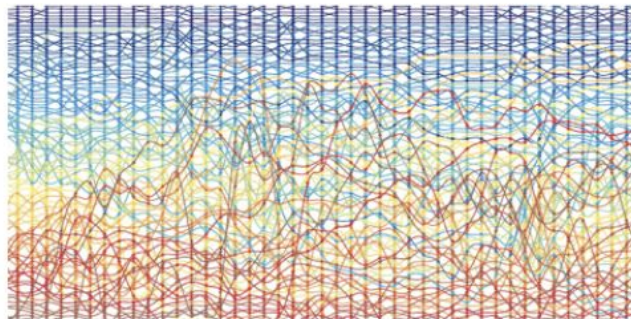
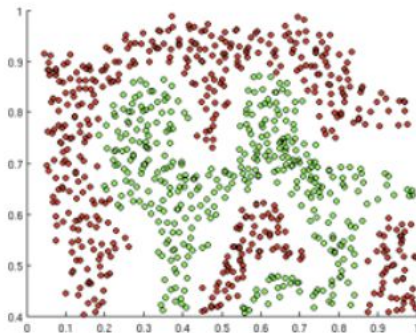
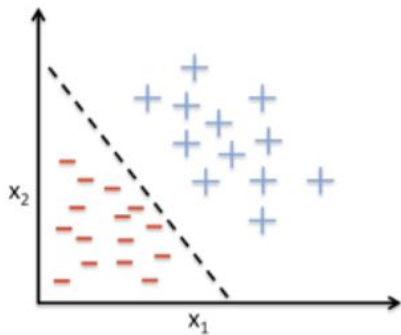
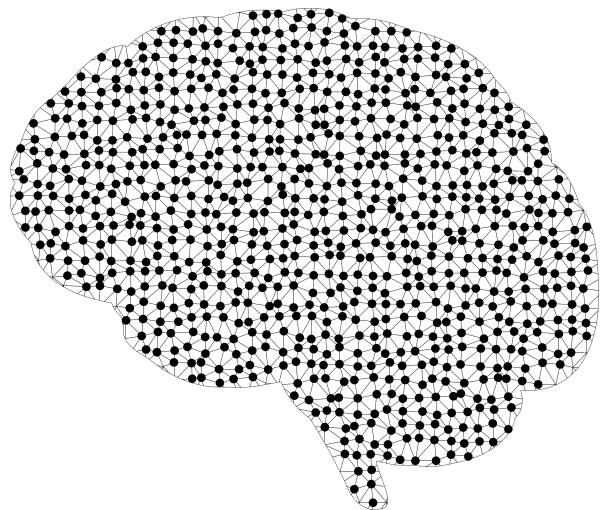
Received: 23 May 2017  
Accepted: 30 August 2017  
Published online: 15 September 2017

Safoora Yousefi<sup>1</sup>, Fatemeh Amrollahi<sup>1</sup>, Mohamed Amgad<sup>1</sup>, Chengliang Dong<sup>2</sup>, Joshua E. Lewis<sup>3</sup>, Congzheng Song<sup>4</sup>, David A. Gutman<sup>5</sup>, Sameer H. Halani<sup>6</sup>, Jose Enrique Velazquez Vega<sup>7</sup>, Daniel J. Brat<sup>7,8</sup> & Lee A. D. Cooper<sup>1,3,8</sup>



# Notes

- Advances in computing
- Availability of large datasets
- Ability to detect highly complex, non-linear patterns
- Ability to predict many outputs at once
- Transfer learning to *transfer* patterns learned by one model to another



# Resources

## INTERFACE

[rsif.royalsocietypublishing.org](http://rsif.royalsocietypublishing.org)

Headline  
review



**Cite this article:** Ching T *et al.* 2018  
Opportunities and obstacles for deep learning  
in biology and medicine. *J. R. Soc. Interface* **15**:  
20170387.  
<http://dx.doi.org/10.1098/rsif.2017.0387>

## Opportunities and obstacles for deep learning in biology and medicine

---

Travers Ching<sup>1,†</sup>, Daniel S. Himmelstein<sup>2</sup>, Brett K. Beaulieu-Jones<sup>3</sup>,  
Alexandr A. Kalinin<sup>4</sup>, Brian T. Do<sup>5</sup>, Gregory P. Way<sup>2</sup>, Enrico Ferrero<sup>6</sup>,  
Paul-Michael Agapow<sup>7</sup>, Michael Zietz<sup>2</sup>, Michael M. Hoffman<sup>8,9,10</sup>, Wei Xie<sup>11</sup>,  
Gail L. Rosen<sup>12</sup>, Benjamin J. Lengerich<sup>13</sup>, Johnny Israeli<sup>14</sup>, Jack Lanchantin<sup>17</sup>,  
Stephen Woloszynek<sup>12</sup>, Anne E. Carpenter<sup>18</sup>, Avanti Shrikumar<sup>15</sup>, Jinbo Xu<sup>19</sup>,  
Evan M. Cofer<sup>20,21</sup>, Christopher A. Lavender<sup>22</sup>, Srinivas C. Turaga<sup>23</sup>,  
Amr M. Alexandari<sup>15</sup>, Zhiyong Lu<sup>24</sup>, David J. Harris<sup>25</sup>, Dave DeCaprio<sup>26</sup>,  
Yanjun Qi<sup>17</sup>, Anshul Kundaje<sup>15,16</sup>, Yifan Peng<sup>24</sup>, Laura K. Wiley<sup>27</sup>,  
Marwin H. S. Segler<sup>28</sup>, Simina M. Boca<sup>29</sup>, S. Joshua Swamidass<sup>30</sup>,  
Austin Huang<sup>31</sup>, Anthony Gitter<sup>32,33</sup> and Casey S. Greene<sup>2</sup>

---

# Resources

- Neural Networks and Deep Learning by Michael Neilson  
[\[http://neuralnetworksanddeeplearning.com/\]](http://neuralnetworksanddeeplearning.com/)
- Deep Learning by Ian Goodfellow and Yoshua Bengio and Aaron Courville  
[\[http://www.deeplearningbook.org/\]](http://www.deeplearningbook.org/)
- DeepLearning.ai on Coursera  
[\[https://www.coursera.org/deeplearning-ai\]](https://www.coursera.org/deeplearning-ai)

# Resources

A mostly complete chart of

## Neural Networks

©2016 Fjodor van Veen - asimovinstitute.org

